CHARACTERISATION OF THE HUMAN CYP2A7 GENE: AN ANALYSIS OF ITS STRUCTURE, REGULATION AND EXPRESSION

Shaohong Ding

Ph.D.

University of Edinburgh

1995

DECLARATION

I declare that this thesis has been completed by myself, and has resulted from my own research. Experiments conducted by other people are appropriately acknowledged.

Acknowledgements

I would like to take this opportunity to express my deep appreciation to Professor Roland Wolf who offered me the chance to work in his laboratory and supervised the research.

My heartfelt thanks also go to the following people for their kind encouragement, valuable comments, suggestions and advice during the final stage of my thesis writing:

Dr. G. J. Moffat, Dr. C. Henderson, Dr. M. Pritchard, Dr. M. Paine and Dr. T. Friedberg.

I would like to thank Dr. B. G. Lake for carrying out the HPLC assay of coumarin-7-hydroxylase activity, Dr. D. Kioussis for allowing me to use the human genomic cosmid library and Dr. S. Keyse for kindly donating the human skin fibroblast cell line (FEK4). Moreover, I would like to acknowledge the support given by all the members of the Molecular Pharmacology Unit of the Imperial Cancer Research Fund (ICRF).

Finally I would like to thank Xiaoquan for all that she has done for me.

Publications Arising from the Research

1. S. Ding, B. G. Lake, T. Friedberg and C. R. Wolf (1995), Expression and alternative splicing of cytochrome P450, CYP2A7. *Biochem. J.*, **306**, 161-166.

2. S. Ding, and C. R. Wolf. Characterisation of alleles of the human cytochrome P4502A7 gene. Manuscript in preparation.

3. Ding, S-H. and Wolf, C.R. (1992), Characterisation of *CYP2A6* and *CYP2A7* expression in huamn liver using PCR. *J. of Basic and Clinic. Physiol. & Pharmac.*, **3**, 149, (Special Issue, Proceedings of the 9th International Symposium on Microsomes and Drug Oxidations).

4. Ding, S-H. and Wolf, C.R. (1992) Cloning and characterisation of human *CYP2A7* gene. *J. of Basic and Clinic. Physiol. & Pharmac.*, **3**, 109, (Special Issue, Proceedings of the 9th International Symposium on Microsomes and Drug Oxidations).

Abbreviations

A	Adenine
Ah	Aromatic hydrocarbon
APS	Ammonium persulphate
ATP	Adenosine triphosphate
ß-NF	ß-naphthoflavone
bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
cAMP	Cyclic adenosine monophosphate
CAT	Cloramphenicol acetyl transferase
cDNA	Complementary DNA
CIP	Calf intestine phosphatase
Coh	Coumarin hydroxylase
CsCl	Caesium chloride
CYP	Cytochrome P-450
DBP	Albumin gene D region binding protein
DEX	Dexamethasone
dH_2O	Deionised water
ddH_2O	Distilled deionised water
DMEM	Dulbecco's modification of Eagle's medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ds	Double stranded
E. coli	Escherichia coli
EDTA	Ethylene diamine tetraacetic acid
EtBr	Ethidium bromide
FCS	Foetal calf serum
G	Guanine
GRE	Glucocorticoid responsive element
HPLC	High pressure liquid chromatograghy
HRP	Horseradish peroxidase
hr	Hour(s)

IgG	Immunoglobulin G
IPTG	Isopropyl-1-thio-B-D-galactoside
kb	Kilobase pairs
kDa	Kilodaltons
min	Minute(s)
MOPS	3-(N-morpholino)propane sulphonic acid
mRNA	Messenger ribonucleic acid
NADPH	ß-nicotinamide adenine dinucleotide phosphate (reduced form)
NDEA	N-nitrosodiethylamine
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
OD	Optical density
P450	Cytochrome P450
PB	Phenobarbital
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
RFLP	restriction fragment length polymorphism
RNA	Ribonucleic acid
rpm	Revolutions per minute
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecylsulphate-polyacrylamide gel electrophoresis
ss	Single stranded
SSC	150 mM sodium chloride, 15mM sodium citrate
T	Thymidine
TBE	Tris-boric acid-EDTA
TBST	Tris buffered saline with 0.5% Tween 20
TCDD	Tetrachlorodibenzo-p-dioxin
TCPOBOP	1,4 bis 2-(3,5 dichloropyridyloxy) benzene
TEMED	N,N,N',N',-tetra-methylethylenediamine
Tris	Tris (hydroxymethyl)amino ethane
UV	Ultraviolet
X-gal	5-bromo-4-chloro-3-indocyl-β-D-galactoside
XRE	Xenobiotic response element

Table of Contents

ABSTRACT

Two genomic clones, CoIIA and LIIA, encoding a protein highly homologous to CYP2A7, were isolated. The clone CoIIA, isolated from a human cosmid library, contained a full length version of the *CYP2A7* gene. The sequence comparison indicated that there were thirteen base pair differences resulting in five amino acid changes between the new gene and *CYP2A7*, suggesting that the gene was an allele of *CYP2A7*, designated *CYP2A7A*. The gene was about 8 kb long and contained 9 exons encoding a protein of 494 amino acids. The 0.5 kb 5' flanking region of *CYP2A7A* contained several putative promoter elements including a typical TATA box, a steroid regulatory element (SRE), and a HepG2-specific factor-1 binding sequence (HPF-1). The latter played an essential role in the expression of *CYP2A7A* in HepG2 cells. However, the function of the SRE element is still unclear.

The restriction map and partial sequences of the second clone LIIA, isolated from a human genomic DNA library EMBL3, showed that this gene was identical with neither *CYP2A7* nor *CYP2A7A*. The results indicated that it was another allele of *CYP2A7*, designated *CYP2A7B*. Of the 18 individuals tested using PCR/restriction fragment length polymorphism (RFLP), the frequency of *CYP2A7B* homozygotes was approximately 44.4%; that of *CYP2A7A* homozygotes was 33.3% and the frequency of heterozygotes of *CYP2A7A /CYP2A7B* was 22.2%.

In order to investigate the relationship between the expression levels of human *CYP2A* alleles and the polymorphism of coumarin hydroxylase activity in man, three cDNAs, *CYP2A6*, *CYP2A7* and an alternatively spliced version of *CYP2A7* (*CYP2A7AS*) were cloned. The last one missed exon 2 but contained a 10 bp segment of intron 1. Translation of *CYP2A7AS* resulted in an in frame deletion of 51 amino acids. The expression of these cDNAs in COS-7 cells showed that both *CYP2A6* and *CYP2A7* generated a protein of 49 kDa, while the protein product of *CYP2A7AS* was about 44

kDa. Only CYP2A6 protein had the coumarin hydroxylase activity. The expressions of CYP2A6 and CYP2A7 mRNAs were established in six human liver samples with RT-PCR (reverse transcription followed by PCR) and *Pst*I digestion. The relative level of CYP2A6 to CYP2A7 was found to range from 1:0.5 to 1:3. The expressions of CYP2A7 and CYP2A7AS mRNA were also investigated. In one of the five liver RNAs studied, the alternatively spliced CYP2A7 mRNA was 3 to 4-fold more abundant than the normal mRNA. The other samples contained very low levels of this mRNA. Interestingly, CYP2A7AS was the major CYP2A7 mRNA detected in a human skin fibroblast cell line. These data have supported the previous findings that alternative splicing may be an important factor in determining the levels of many human cytochrome P450s.

CHAPTER 1: INTRODUCTION

1.1 Metabolism of Xenobiotic and Cytochrome P450: General View and Historical Background

Xenobiotic are compounds foreign to the body, including drugs, environmental pollutants, industrial chemicals and plant metabolites. Many of these compounds are lipophilic and readily absorbed across cell membranes. In mammals, however, these nonpolar compounds cannot be excreted unless they are metabolised to more polar products. Xenobiotic-metabolising enzymes in mammals are responsible for the detoxification and excretion of these foreign chemicals (Schenkman, 1993). By a range of chemical reactions these enzymes convert lipophilic chemicals into highly water-soluble polar compounds that can be eliminated. The metabolic reactions of xenobiotic are normally divided into phase I (functionalisation) and phase II (conjugation) (Table 1.1). The main function of phase I reactions is the creation of a chemically reactive functional group, such as OH, -NH_2, -SH and -COOH, in the substrate. Phase II enzymes can then use these functional groups to yield hydrophilic products. The phase I enzymes include, among others, cytochrome P450 (P450). P450s are very important in the oxidative, peroxidative, and reductive metabolism of numerous xenobiotic as well as endogenous compounds such as steroid hormones, bile acids, fatty acids, prostaglandins and their derivatives (Zimniak and Waxman, 1993). With certain compounds, however, the intermediates of oxygenation reactions catalysed by P450, such as epoxides, can attack the cellular biomolecules DNA, RNA, and protein, producing toxicity, cell death, mutation and transformation.

The metabolisms catalysed by P450s share some common features. First, a single P450 enzyme can metabolise numerous structurally diverse chemicals. For example, CYP2D (P4502D, see section 1.2.1) is involved in the metabolism of at least 25 different

Table 1.1 Reactions classed as phase I or phase II metabolism (Gibson and Skett, 1986)

Phase I	Phase II
Oxidation	Glucuronidation/Glucosidation
Reduction	Sulphation
Hydrolysis	Methylation
Hydration	Acetylation
Dethioacetylation	Amino acid conjugation
Isomerization	Glutathione conjugation
	Fatty acid conjugation
	Condensation

chemicals (Gough *et al.*, 1990). Second, a single substrate can be metabolised to varying degrees by several different P450 enzymes. Testosterone can be metabolised at different ring positions. Certain P450s are active toward hydroxylating testosterone in a stereo-specific manner at one or more ring positions, while another form of P450 will hydroxylate this substrate at other positions (Gonzalez, 1989).

P450 has been the subject of intense studies for over 40 years. Since the 1950s, it has become clear (Müller and Miller, 1953) that liver microsomes are the major source of P450. In the 1960s, Hayashi and Nozaki (1969) discussed that the P450s catalysed the incorporation of a single atom of molecular oxygen into the substrate with the concomitant reduction of the other atom to water. After the isolation of a partially purified P450 from phenobarbital-treated rabbit liver microsomes (Lu and Coon, 1968), a number of other forms of P450 were purified from different species, ranging from primitive bacteria to highly developed mammals. Although liver is the major source of P450s, they can be found in almost all other tissues to a variable extent (Arinc, 1993; White *et al.*, 1991; Bergh and Strobel, 1992). P450 has also been found in some cultured cell lines, such as human colon tumour cell line LS174T (Strobel *et al.*, 1993), and human skin cell line (described in Chapter 4).

1.2 Evolution and Nomenclature

1.2.1 Cytochrome P450 nomenclature

Over the last 3 decades, tremendous progress has been made in understanding the P450 system. Since the first P450 enzyme was partially purified in the 1960's, a total of 214 P450 genes and 12 putative pseudogenes, which were isolated and characterised from 31 different eukaryotes and nine prokaryotes, have been listed to the end of October, 1992 (Nelson *et al.*, 1993). The rapidly expanding number of genes within the P450 superfamily and a variety of different nomenclatures used in different laboratories made it necessary to establish a standard nomenclature system of P450 genes.

The name cytochrome P450 was first used to describe a hemoprotein having a major absorption at 450 nm after binding carbon monoxide (Omura and Sato, 1962; 1964). However, the Nomenclature Committee of the International Union of Biochemistry (NCIUB) prefers the term "heme-thiolate protein" instead of "cytochrome" for P450, because P450 proteins are, in fact, not "cytochrome" in the true meaning of this terminology (Palmer and Reedijk, 1989).

It has been known that some P450s have very high substrate specificity, while other P450s have broad substrate specificities. Orthologous P450s may have different substrate specificities and, conversely, nonorthologous P450s or P450s in different subfamilies may have similar substrate specificity, particularly the P450 genes in Family 2. Therefore the nomenclature system cannot be based on P450 catalytic activities or functions.

Based entirely on the homology of complete amino acid sequences, the first nomenclature system was established in 1987 (Nebert *et al.*, 1987) and extended in subsequent updates (Nebert *et al.*, 1989a; 1991, Nelson *et al.*, 1993). This system divides P450s into gene families and subfamilies. Each family has diverged from one another about 600-900 **m**illion **y**ears **a**go (MYA), and the P450 protein sequences within a single family are > 40% similar with a few exceptions, one of which is the CYP2D subfamily. Some P450s in this family are slightly less than 40 % similar to other proteins in the CYP2 family (Gonzalez, 1993). Any two mammalian sequences of proteins within the same subfamily are greater than 55% identical, and they have diverged from one another within 150 million years. All genes within a family have been shown to contain the same number of exons and similar intron-exon boundaries..

Briefly, the nomenclature system includes the points as follows (Nelson *et al.*, 1993)

1) A P450 gene or cDNA is named using the italicised root symbol "***CYP***" ("***Cyp***" for the mouse), denoting ***cy***tochrome ***P***450.

2) An Arabic number designating the P450 family, and a letter indicating the subfamily when two or more subfamilies are known to exist within that family, and another Arabic numeral represents the individual gene. With mouse genes or cDNAs, the final number is generally preceded by a hyphen. "***P***" ("***p***" in mouse) after the gene number is used to denote a pseudogene.

3) It is recommended to non-italicise product of the gene (enzyme) by a similar nomenclature system, for example, '**CYP1A1**" for the mRNA and protein in all species including mouse. The italicised "***CYP1A1***" ("***Cyp1a-1***" in mouse) represents the gene or cDNA. Another possibility for designating the protein might be "P4501A1" or simply "1A1"

Based on these points, *CYP2A6* represents a human P450 gene or cDNA, family 2, the sixth gene in subfamily A; CYP2A6 is the enzyme or mRNA encoded by the *CYP2A6* gene. *Cyp2a-4* represents a mouse P450 gene or cDNA, family 2, the fourth gene in subfamily A, and Cyp2a4 is the enzyme or mRNA of the *Cyp2a-4* gene.

Although the use of this nomenclature system is recommended, the trivial names can still be used. For the protein, there is no hyphen in "P450", no Greek letters, and no subscripts or superscripts. For example, $P450_{7\partial}$, $P450_{17\partial}$, and $P450_{arom}$ should be referred to as P450c7, P450c17 and P450arom, respectively.

1.2.2 Evolution of CYP superfamily

P450s are presumed to have been present in the earliest organisms, and all the existing species of P450 are derived from a common ancestor more than 3 billion years ago (Nelson *et al.*, 1993). During the evolutionary process the P450 system has diverged and duplicated to a superfamily, being found throughout the animal and plant kingdoms. An evolution tree (Figure 1.1) for the *CYP* gene superfamily was generated by comparing amino acid sequence data (Nebert and Gonzalez, 1987; Nelson and

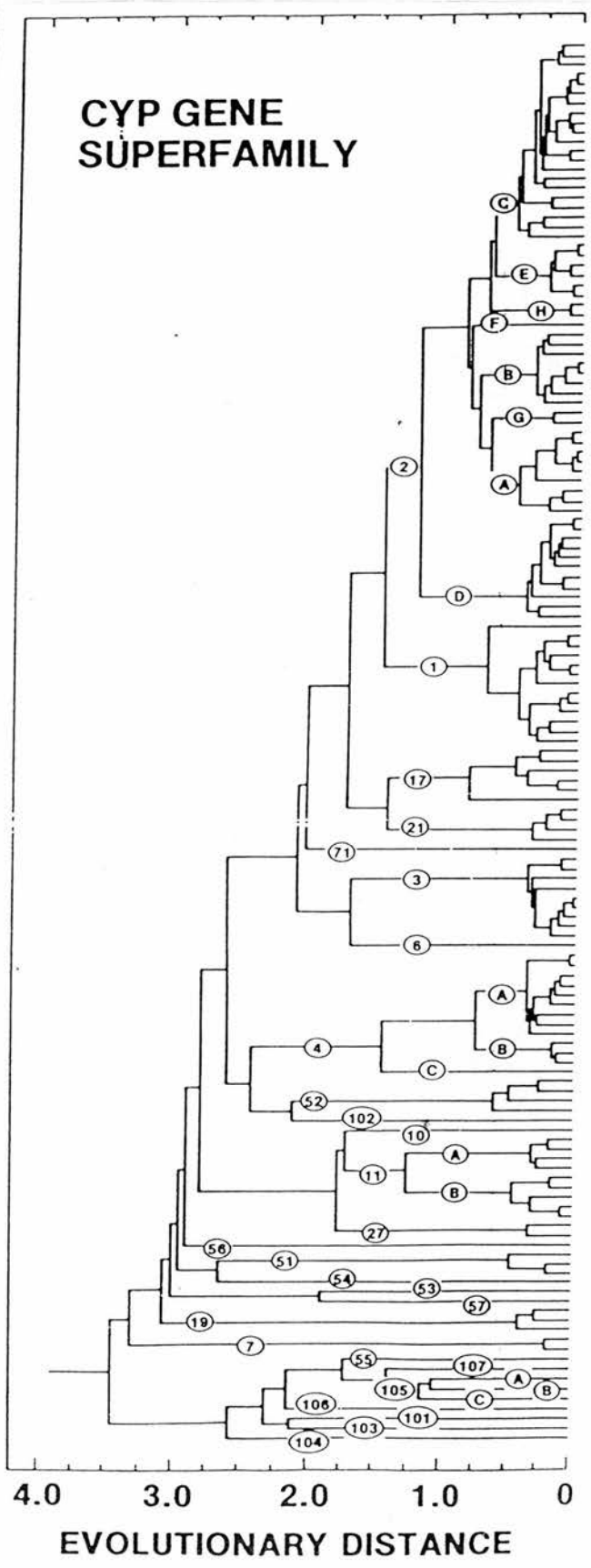


Figure 1.1 Unweighted-pair-group method of analysis (UPGMA) of the P450 superfamily (Nebert *et al.*, 1991). The genes within each family (denoted by numbers) represent all those that have been characterised in various species. The divergence between bacterial and eukaryotic genes (evolutionary distance = 2.5) has been set at 1400 million years-base pairs (Mybp). Estimates of branching in the oldest portion of the tree are subject to the largest error, and molecularly driven events during evolution contribute to the uncertainty of UPGMA branching patterns. This tree was calculated by using amino acid sequences deduced from 147 of the 154 cDNA sequences available.

Strobel, 1987). Using the species divergence time generated from fossil evidence and the amino acid differences between a P450 protein in two species, the evolutionary distance and unit evolutionary period (UEP, the time in millions of years required for a 1% change in amino acid sequence) can be calculated. The UEPs of P450s range from 2.3 to 4.2 (Nelson and Strobel, 1987), suggesting that P450 genes are rapidly diverging.

The earliest P450s are those that metabolise steroids and fatty acids; for example, the fatty acid metabolising *CYP4* and the steroid inducible *CYP3* genes diverged more than 1 billion years ago. Then the *CYP1* and *CYP2* gene families formed about 800 MYA followed by the expansion of the *CYP2* gene family to eight subfamilies at approximately 400 to 600 MYA. To date, a minimum of 54 genes have been found in this family (Henderson and Wolf, 1992).

The 'explosion' in the number of new *CYP2* genes may be related to the emergence of mammals onto land several millions years after plants were established and most of these genes are believed to represent a 'surviving warfare' between animal and plant (Gonzalez and Nebert, 1990). After animals and plants diverged approximately 1.2 billion years ago, animals began to ingest plants. As a means of defence, plants countered by developing new stress metabolites to make them less palatable and digestible; then animals responded with new P450 genes. The presence of detoxifying enzymes encoded by these P450 genes allowed animals to survive in their new environment.

1.2.3 Evolution of CYP2A subfamily

During the process of evolution, a tremendous expansion in the *CYP2* gene family has occurred within the past 400 to 800 million years which resulted in a diverging of eight subfamilies. This diversification has been suggested to coincide with emergence of vertebrates on to land. These animals could only survive on plant diets if they had

developed their own specialised battery of P450s, particular those in CYP2 family (Gonzalez, 1992). CYP2 family is composed of ten subfamilies with different members (Nelson *et al.*, 1993). The enzymes within these subfamilies are involved in the metabolism of many chemicals including drugs and procarcinogens, and the enzyme activities are under distinct inducer-dependent and developmental control (Squres and Negishi, 1988; Matsunaga *et al.*, 1988).

About 75 - 80 million years ago, a duplication event occurred, generated the *CYP2A3* gene and the precursor to the *CYP2A1* and *CYP2A2* genes in rat, followed by another diverging event approximately 25 million years ago, which resulted in the *CYP2A1* and *CYP2A2* genes (Gonzalez, 1989). The amino acid of the CYP2A3 exhibits 71% and 73% similarity to that of the CYP2A1 and CYP2A2 proteins, respectively, while the CYP2A2 shares 88% of the amino acid sequence of CYP2A1 protein. In mouse, two members of *Cyp2a* gene subfamily, designated *Cyp2a-4* and *Cyp2a-5,* have been discovered and the gene clusters are located on mouse chromosome 7. The deduced amino acid sequences of both genes are 98% similar, indicating that they diverged about 4 to 12 million years ago. In spite of their highly similar sequences, *Cyp2a-4* encodes a 15α-hydroxylase, and *Cyp2a-5* encodes a coumarin 7- hydroxylase. Both mouse *Cyp2a* genes exhibit 70% and 75% deduced amino acid sequence similarity with rat CYP2A1 and CYP2A2, respectively, but 90% similarity with rat CYP2A3. Therefore, mouse genes have been designated as orthologous to the rat *CYP2A3* gene. Two human *CYP2A* cDNAs, designated *CYP2A6* and *CYP2A7*, have been isolated and located on chromosome 19, between 19q12 and 19q13.2 (Miles *et al.*, 1988). The deduced amino acid sequence of human *CYP2A* gene is 85% similar to CYP2A3, 69% and 65% similar to CYP2A1 and CYP2A2, respectively.

Although the gene numbers of *CYP2A* subfamily are variable in different species, they exhibit a single conserved activity, namely coumarin 7-hydroxylation. However, the

substrate turnover in different mammals and strains is varied (Gonzalez, 1992; Lindberg *et al.*, 1989 and 1992).

1.2.4 The molecular mechanism of cytochrome P450 evolution

Gene duplication produces two identical copies and these may retain their original function, producing the RNA species or protein. Alternatively, one of the copies may diverge by the occurrence of molecular events at the DNA level, such as point mutation, deletion and insertion, to form a functionless pseudogene. More importantly, the gene duplication may result in the emergence of a new gene with markedly different function (Li and Graur, 1991). If a new gene confers an evolutionary advantage(s), it can become fixed under selective pressures, and can ultimately be spread throughout a population (Nebert and Gonzalez, 1987; Nelson and Strobel, 1987). This is considered as part of a process of molecular drive (Gonzalez and Nebert, 1990). Figure 1.2 represents schemes of *CYP* gene duplication.

One of the mechanisms involved in molecular drive is gene conversion. This event takes place during gene duplication in which some portion of a gene is replaced by the corresponding part of a nearby gene. Several reports have presented evidence for gene conversion between closely linked *CYP* genes, for example within *CYP2A* and within *CYP2D* genes (Figure 1.2) (Matsunaga *et al.*, 1988; Gonzalez and Nebert, 1990), therefore gene conversion is hypothesised as a major determinant in P450 evolution (Nebert and Gonzalez, 1987). However based on the sequence comparison within the rat *CYP2D* gene cluster (Matsunaga *et al.*, 1990), in which several independent events homogenised nucleotide sequences surrounding the site coding the heme binding region, Gotoh (1993) argued that gene conversion appeared to be conservative rather than progressive in nature, and was probably caused by a repair mechanism against accidental gene disruptions. These results question whether gene conversion played a

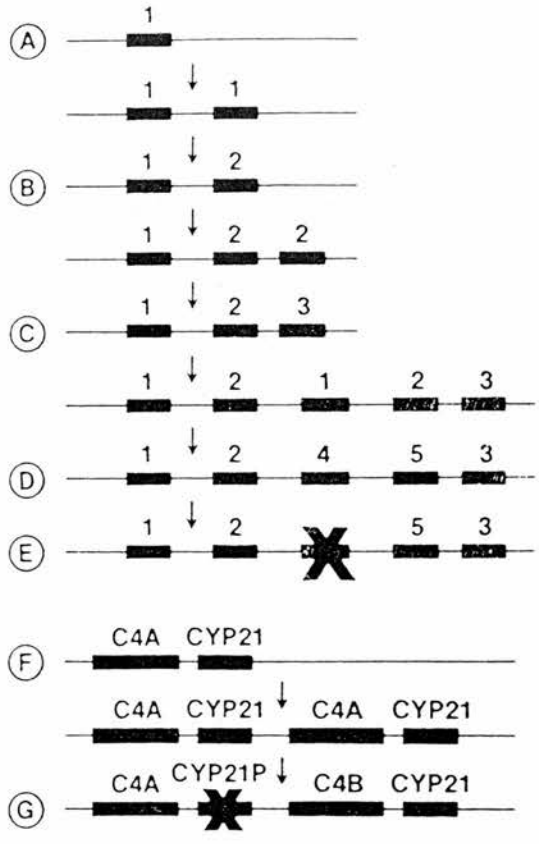


Figure 1.2 Schemes showing gene duplication and divergence events during evolution. The black boxes represent functional genes in a subfamily along the chromosome (denoted by line). Transitions of gene 1 to two gene 1's (A) or of gene 2 to two gene 2's (B) are examples of duplication events. Formation of gene 2 from gene 1 (A and B) or of gene 4 and 5 from gene 1 and 2, respectively, (C and D), represent examples of divergence events. It is believed that the original human complement 4 gene (C4A) and *CYP21* gene (White *et al.*, 1985), existed (F) as a single gene 150 million years ago, and constitutes an example in which both genes duplicated together and occur in tandem in present-day humans (G). The X represents a pseudogene that has formed as the result of mutation, deletion, or crossover. This figure is adapted from Nebert *et al.*, 1989b.

central role in the progressive evolution of P450. Taken together, the evolution of P450, whatever the mechanism, appears to have been governed by a more complicated process than previously thought.

1.3 Structure and Metabolic Reactions of Cytochrome P450

1.3.1 Structure of cytochrome P450

Cytochrome P450 has been classified as a heme-containing enzyme with a single iron protoporphyrin IX as the prosthetic group. Dioxygen is bound, reduced and activated at this site (figure 1.3, A). The amino acids surrounding the heme influence the spectral characteristics of the heme resulting in a typical P450 absorption band at around 450 nm once the ferric iron has been reduced and carbon monoxide has bound (Klingenberg, 1958). These enzymes exist in multiple forms with a molecular weight of the monomers of approximately 45,000 - 55,000 (450-500 amino acid residues). Several lines of evidence suggest that the P450 protein is tightly bound in the membrane with its N-terminal (amino terminus) peptide serving as an anchor, and the major part of the polypeptide chain is exposed to the cytoplasmic surface of the endoplasmic reticulum to probably form a globular structure. The heme iron is parallel to the endoplasmic reticulum membrane (Figure 1.3B).

At present, more than 160 primary structures of P450 protein have been characterised and structural homology has been observed in some peptide regions among all of these enzymes; for example, the heme-binding Cys residue near the C-terminus (carboxy terminus) is invariant. The highly hydrophobic signal/anchor segment at the N-terminus is followed by a short positively charged sequence and a proline rich cluster; the remainder of the sequence exhibits alternating hydrophobic and hydrophilic character (Black, 1993).

Figure 1.3A: Structure of ferric protoporphyrin IX, the prosthetic group of cytochrome P450 (Adapted from Gibson and Skett, 1986).

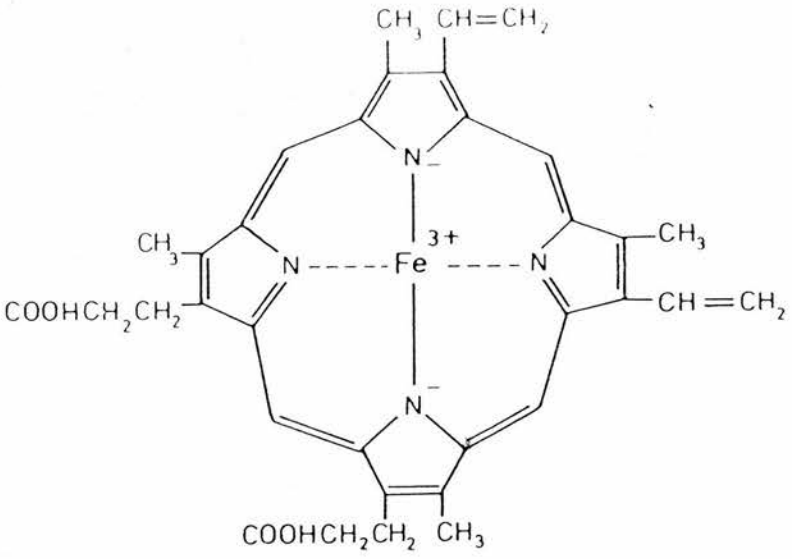
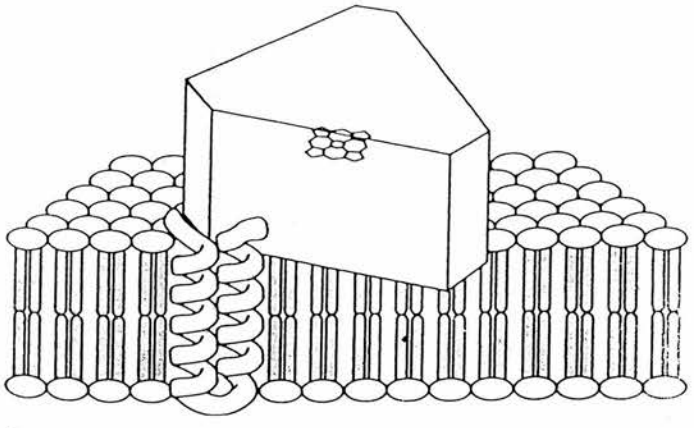


Figure 1.3B: A model of microsomal cytochrome P450 structure (Nelson and Strobel, 1988).



Alignment of primary structure with X-ray crystallographic data has been used to predict five substrate-binding domains (Gotoh and Fujii-Kuriyama, 1989). Based on theoretical calculations, P450s contain α-helix as well as ß-sheet domains. Experimentally, however, the tertiary structure is only known for P450 101 ($P450_{cam}$) from the bacterium *Pseudomonas putida* based on X-Ray analysis (Figure 1.4). Although the three-dimensional structure is not yet known for any eukaryotic cytochrome P450, many advances have been made to elucidate it, especially regarding the membrane topology (Black, 1992). By alignment of 34 sequences of microsomal P450, four to eight potential trans-membrane regions have been demonstrated (Nelson and Strobel, 1988). However, recent experimental results showed that eukaryotic P450s are anchored in the membrane by only one or two trans-membrane peptides located at the N-terminal end, leaving the globular, "P450 101-like" domain outside the membrane (Black, 1992).

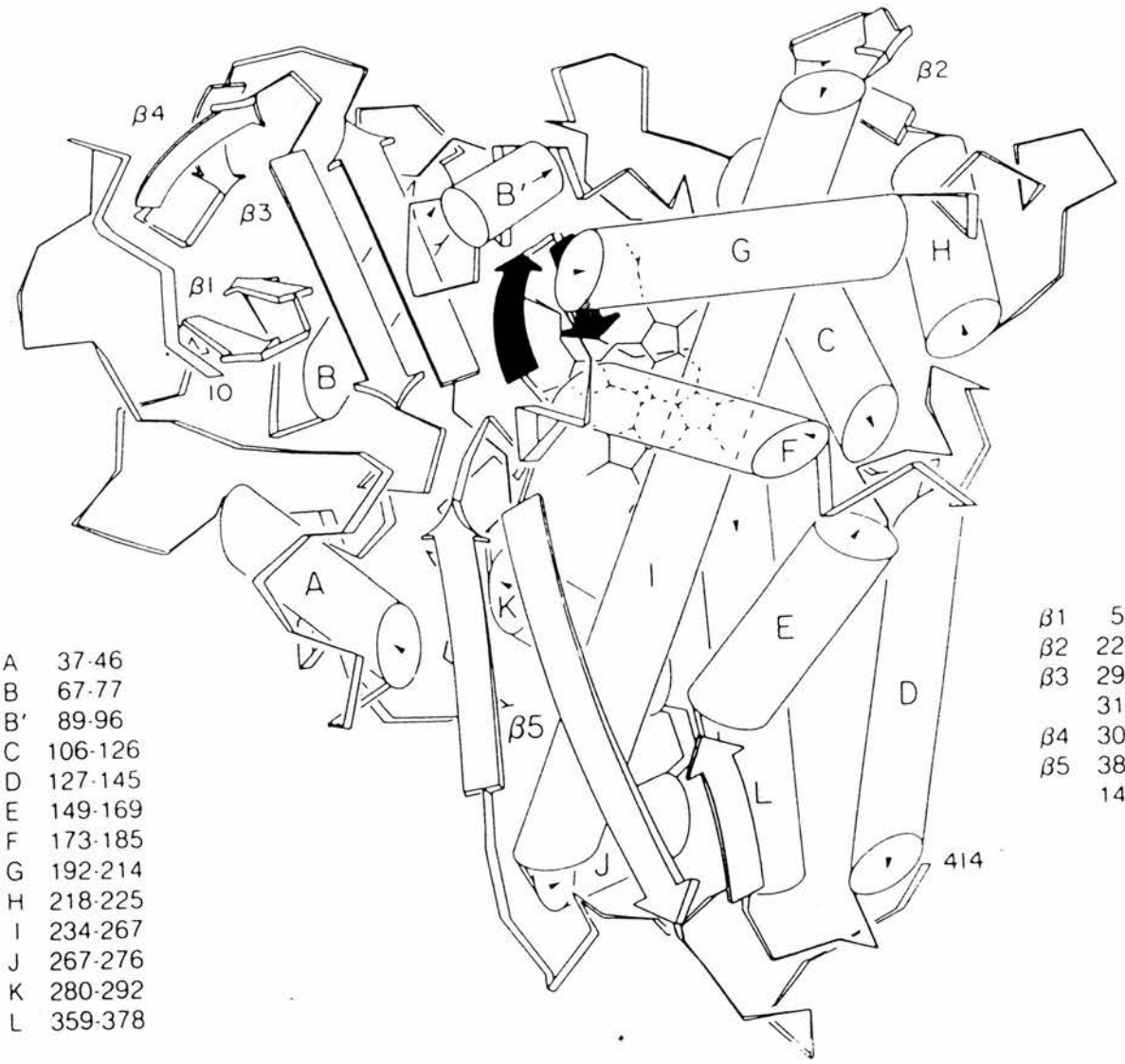
1.3.2 Reactions catalysed by cytochrome P450

It is known that humans are exposed to more than 200,000 man-made chemicals or environmental pollutants in their daily lives (Porter and Coon, 1991). Many of these pollutants are believed to be potential substrates for P450, and many may also serve as inducers or inhibitors of different P450 enzymes. Among the huge number of structurally diverse chemicals, the only common feature appears to be a degree of lipophilicity. The overall reaction is given in equation 1. It is in accord with findings in a number of laboratories and with the known stoichiometry of the hydroxylation reaction.

Equation 1:

$$NADPH + H^+ + O_2 + RH \xrightarrow{\text{Cytochrome P450}} NADP^+ + H_2O + ROH$$

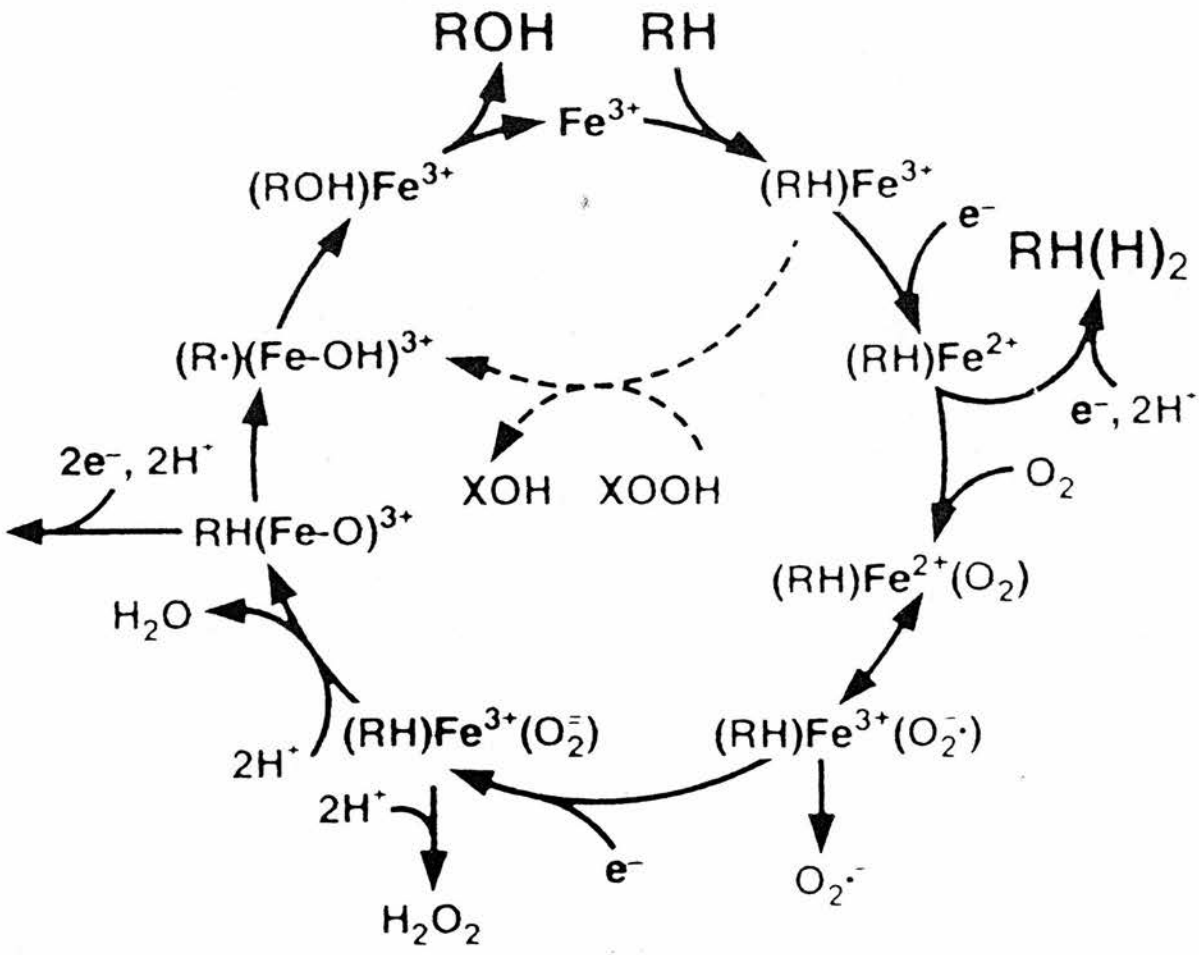
Figure 1.4 High-resolution crystal structure of cytochrome $P450_{cam}$ (Poulos, 1991). Helices are indicated by bars and ß structure by arrows. The shaded region highlights the antiparallel ß pair and ß bulge that contains the axial heme ligand, Cys-357.



Where RH represents an oxidizable substrate and ROH the hydroxylated metabolite. During the reaction, reducing equivalents derived from NADPH + H^+ are consumed, and one atom of molecular oxygen is incorporated into the substrate whereas the other oxygen atom is reduced to water. In an enzyme-free system, two extremely high activation energy barriers have to be surmounted during the reaction: the dissociation of the dioxygen bond (460 kJ/mol) and the dissociation of the carbon hydrogen bond of the substrate (420 kJ/mol) (Jung and Ristau, 1978). However, the reaction catalysed by P450 can proceed with an activation energy of only 40-70 kJ/mol (Rein and Jung, 1993).

Figure 1.5 is a more detailed scheme of the reaction cycle catalysed by P450, which is modified from an earlier version (White and Coon, 1980) The feature of the reaction is the ability of the heme iron to undergo cyclic oxidation/reduction reactions in conjunction with substrate binding and oxygen activation. It may be considered in two steps. The first step in the reaction cycle is substrate binding, which perturbs the spin state equilibrium of the cytochrome and facilitates uptake of the first electron. Substrates that undergo reduction rather than oxygenation, such as epoxides, N-oxides, nitro and azo compounds, and lipid hydroperoxides, accept two electrons to give $RH(H)_2$ (Porter and Coon, 1991; Kominami, 1993). To initiate the oxidative reactions, O_2 is bound to the ferrous P450 with co-ordination to iron trans to thiolate. A substrate binding-induced shift in mid-point redox potential of cytochrome P450 protein to a more positive value results in a greater electromotive force for subsequent facile electron transfer of the second electron (For review see Schenkman *et al.*, 1982). The next step is oxygen insertion and product release. The precise oxidation states of iron and oxygen in this intermediate is far from being clear but involve splitting the oxygen-oxygen bond with the uptake of two protons at some stage, and the reaction of an activated oxygen and the release of H_2O. The Fe^{2+}-O_2/substrate complex is unstable,

Figure 1.5. Catalytic cycle of cytochrome P450. Fe represents the heme iron atom at the active site, RH the substrate, $RH(H)_2$ a reduction product, ROH the corresponding hydroxylated metabolite, and XOOH a peroxy compound that serves as an alternative oxygen donor (Adapted from Porter and Coon , 1991).



but has been characterised both with the bacterial P450 101 ($P450_{cam}$) and mammalian P450s (Oprian *et al.*, 1983; Guengerich, 1991), and evidence for the other oxygenated complexes has also been seen with P450 enzymes (Blake and Coon, 1989).

Under certain conditions, the oxidation of substrates is not tightly coupled to the electron flow, the catalytic cycle results in the formation of hydrogen peroxide (equation 2), and this uncoupling is dependent on substrate (RH) and its binding to the active site of the heme moiety of P450 (Gorsky *et al.*, 1984; Blanck *et al.*, 1991). Evidence has been accumulated that most of this H_2O_2 is formed from the non enzymatic dismutation of superoxide anion, which is a breakdown product of the $(RH)Fe^{2+}$-O_2 complex (Kuthan and Ullrich, 1982; Figure 1.5).

Equation 2:

$$O_2 + NADPH + H^+ \xrightarrow{\text{P-450}} H_2O_2 + NADP^+$$

On the other hand, it is clear that uncoupling is dependent on the individual P450 involved in the metabolism of a particular chemical. *In vivo*, different substrates lead to a different spectrum of induced P450, thus changing the degree of uncoupling of the specific chemical. The best examined example in this regard is ethanol, which induces CYP2E (Kappus, 1993). Whereas hydrogen peroxide formation by ethanol in non-induced liver microsomes is low, it is increased in ethanol-induced liver microsomes. The hydrogen peroxide and hydroxyl radicals have been related to alcohol-induced liver damage (Albano *et al.*, 1991). This is probably the cause of reactive oxygen formation *in vitro* and *in vivo* (for more detail about the formation of reactive oxygen species, see Kappus, 1993).

P450s can also bioactivate certain environmental pro-carcinogens to cytotoxic and carcinogenic products. In general, P450 enzymes use molecular oxygen in the detoxification of foreign compounds. However, P450-mediated oxidation, at carbon

atoms to form epoxides and at nitrogen and sulphur atoms, can give rise cytotoxic and mutagenic products. The vast majority of environmental procarcinogens including polycyclic aromatic hydrocarbons, nitrosamines and aromatic amines are converted to their ultimate carcinogenic species by these reactions (Wolf, 1986).

Equation 3:

$$R' = R + O_2 + NADPH + H^+ \longrightarrow R - R'(O) + H_2O + NADP^+$$

A summary of these reactions is that P450 can play three distinct roles in the metabolism of xenobiotic and endogenous compounds. First, the enzyme system is a dominant route converting a lipophilic foreign chemical into a more water soluble product, thereby facilitating excretion from the organism, this process is considered to be the detoxification. Second, it can convert certain xenobiotic to more toxic products. Many carcinogens are metabolically activated by P450s to ultimate carcinogenic metabolites. Last, this enzyme system can also metabolise numerous endogenous compounds including steroids, fatty acids, vitamins, and bile acids (Ying and Lu, 1987).

1.4 Roles of Cytochrome P450s in Activation of Chemical Carcinogens and Genetic Susceptibility to Diseases

1.4.1 Activation of procarcinogens by P450

In general, P450s use molecular oxygen in the detoxification of foreign compounds. As an unfortunate consequence of these reactions some procarcinogens are activated to their ultimate carcinogenic forms (Harris, 1991; Guengerich, 1988; Kawajiri *et al.*,

1990). Short-term mutagenicity testing has clearly shown that P450-containing preparations in *Salmonella typhimurium*-based assays can convert the great majority of pre-mutagens into mutagens (Ashby and Tennant, 1991). Since most of these mutagens are also carcinogens, similar metabolic activation processes should occur *in vivo*, too.

Studies involving the metabolism of chemical carcinogens were important in the initial characterisation of P450. Early evidence revealed that NADPH-dependent microsomal enzymes metabolised azo dyes, and that administration of many different chemicals to animals could alter the metabolism of carcinogens (Guengerich, 1988). Since then many studies regarding the catalytic specificities of the P450 enzymes of rats, mice and other experimental animals, as well as humans have been published. Table 1.2 lists some carcinogens which can be activated by P450.

Aflatoxins comprise a structurally-related group of mycotoxin metabolites which are present in a variety of human foodstuffs (Groopman *et al.*, 1988; Wogan, 1973a). Aflatoxin B_1 has been found to cause hepatocarcinogenesis in animals (Wogan, 1973b; Vesselinovitch *et al.*, 1972). Furthermore, there is epidemiological evidence that humans may be susceptible to aflatoxin-induced hepatocarcinogenesis (Peers *et al.*, 1987). Aflatoxin B_1 is considered to be the most carcinogenic of these toxins and requires oxidation of the 8,9 double bond to yield the biologically active aflatoxin B_1-8,9-epoxide which can react with DNA. In addition to the 8,9-epoxide, P450-mediated aflatoxin B_1 oxidation also produces several hydroxylated metabolites including aflatoxin M_1, aflatoxin P_1, and aflatoxin Q_1. The metabolism of aflatoxin B_1 is illustrated in Fig 1.6. In humans, the enzymes in subfamilies CYP3A, CYP2A and CYP2B have been reported to be involved in the bioactivation of aflatoxin B_1 (Aoyama *et al.*, 1990), and the reactions involved are epoxidation, hydroxylation and demethylation (Figure 1.6).

Table 1.2. Association of cytochrome P450 and carcinogen activation (Adapted from Yamazoe and Kato, 1993).

P-450 family	Type of carcinogen oxidation
1A1	Extrahepatic oxidation of arenes (polycyclic aromatic hydrocarbons) and other carcinogens.
1A2	N-Oxidation of arylamines and arylamides
2A	Epoxidation of mycotoxins such as aflatoxin B_1.
2B	Oxidation of long alkyl-chain nitrosamines
2C	Cytclophosphamide 4-hydroxylation. Basal level of hepatic oxidation of arenes.
2E	α-Oxidation of short alkyl chain nitrosamines and benzene oxidation
3A	Epoxidation of mycotoxins such as aflatoxin B_1. Oxidation of arenes
4A	Azo dye reduction
4B	Arylamine N-oxidation

Nitrosamines are a potent group of environmental mutagens and carcinogens. For example, nitrosodimethylamine (NDMA) has been shown to be both an acute hepatotoxin and a potent carcinogen in many animal species (Magee and Barnes, 1967: Druckrey *et al.*, 1969). The tobacco-specific nitrosamines, nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and N'-nitrosonornicotine have been suggested to play a role in human tobacco-related cancer (Hecht *et al.*, 1983). Metabolic activation of NNK involves the α-Hydroxylation of either the methylene or methyl carbon of NNK, resulting in the methylation or pyridyloxobutylation of DNA, respectively (Figure 1.7) (Hecht *et al.*, 1988). CYP2A6, CYP2E1, CYP2B1 and CYP1A2 are effective in the activation of NNK to a mutagen (Smith-TJ *et al.*, 1992).

A therapeutic but carcinogenic agent, cyclophosphamide, must first undergo bioactivation by P450 to manifest alkylating activity. The initial reaction, cyclophosphamide 4-hydroxylation, is catalysed mainly by CYP2C6 and 2C11 in rat liver (Clarke and Waxman, 1989).

1.4.2 Genetic susceptibility to disease

Xenobiotic compounds, including drugs and carcinogens, are metabolised by P450s and other drug-metabolising enzymes. However, there is considerable evidence for inter-individual variation of cytochrome P450 enzyme activities in humans, and these differences may be either inherited or acquired. Consequently, the metabolic oxidation of drugs and other alien chemicals can be highly variable from person to person. For example, some patients may respond better than others to the same drug treatment. This in part makes it difficult to produce a drug appropriate to everyone. In addition, although cigarette smoking is considered to correlate with lung cancer, not every smoker will definitely develop cancer because of a differential susceptibility to tobacco smoke (Bascom, 1991). Therefore, the individual differences in the level of drug-

Figure 1.6. P450-mediated metabolism of aflatoxin B_1 (adapted from Yamazoe and Kato, 1993).

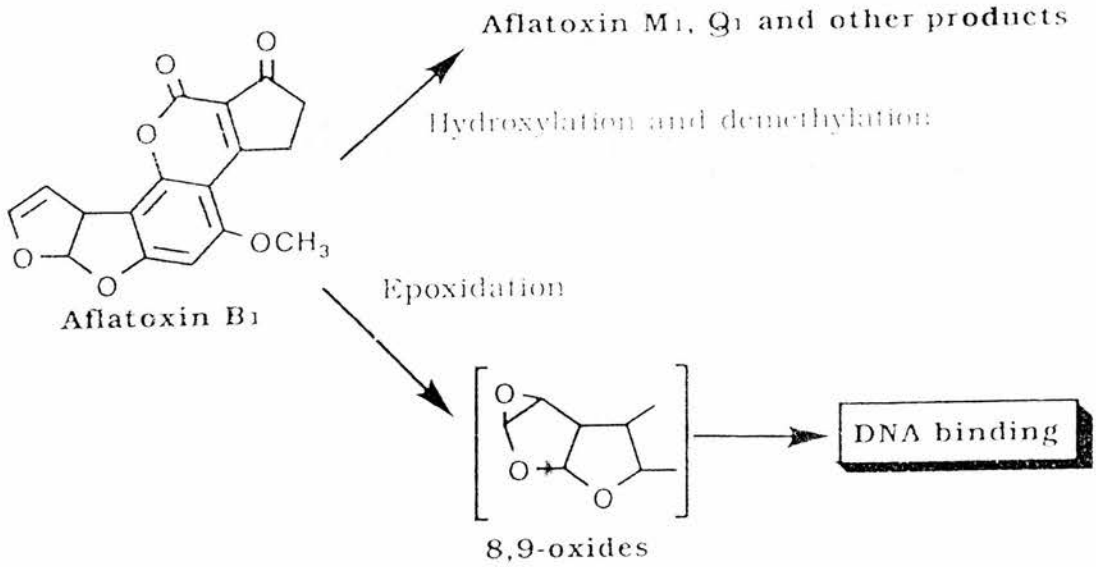
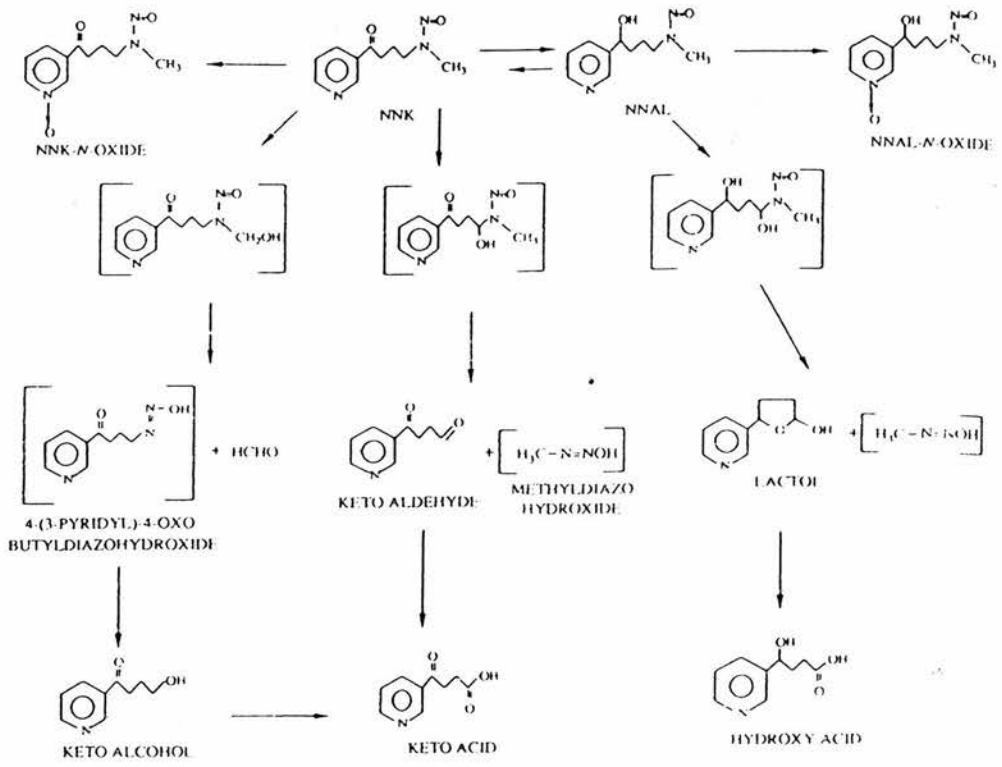


Figure 1.7 Metabolic pathway of NNK. Structures in brackets, hypothetical intermediates (adapted from Smith-TJ *et al.*, 1992)



metabolising enzyme activity may well be a factor in cancer susceptibility and in environmentally linked disease. The understanding of P450s could supply important information for medical treatment and cancer prevention.

In the past 15 years, tremendous progress has been made in the identification of polymorphisms in P450 genes. By studying the metabolic phenotype, it has been found that 5-10% of individuals from different racial groups cannot metabolise the antihypertensive agent debrisoquine (Alvan *et al.*, 1990). Levels of debrisoquine and the 4-hydroxyl metabolite can be measured and used to calculate a metabolic ratio (defined as debrisoquine/4-hydroxyl metabolite). Individuals with a metabolic ratio of greater than 12.6 are classified as poor metabolisers of debrisoquine (PM) and others as extensive metabolisers (EM) (Daly and Idle, 1993). Further studies indicate that CYP2D1 (rat) and CYP2D6 (human) have high debrisoquine hydroxylase activity (Larrey *et al.*, 1984; Brosen and Gram, 1989), and this enzyme is absent from the liver of poor metabolisers (Zanger *et al.*, 1988). In addition, CYP2D is involved in the metabolism of more than 30 clinically important drugs, including antidepressants, neuroleptics, opioids and cardiovascular drugs.

Cloning and sequencing of *CYP2D* genes have assisted in the identification of mutations which give rise to the debrisoquine polymorphisms in rat and human. The results suggest that mutations in the *CYP2D6* gene are common in poor metabolisers and may be responsible for the absence of the CYP2D6 protein. To date, more than 90% of the mutations of *CYP2D6* have been identified (Skoda *et al.*, 1988; Kagimoto *et al.*, 1990; Gough *et al.*, 1990). The most common mutant alleles are characterised by a point mutation or a 1-bp deletion at a splice site recognition sequence that lead to a frame shift.

The debrisoquine hydroxylase phenotype has important consequences both with regard to response to drug therapy and exposure to xenobiotic. Both extensive and poor metabolisers may suffer adverse effects due to low serum concentrations or toxic

concentrations of xenobiotic, respectively (Broken and Gram, 1989). A well studied case is the association between poor metaboliser status and susceptibility to Parkinson's disease (Smith *et al.*, 1992). It has been postulated that this disease may occur as a result of environmental exposure to pesticides similar to MPP^+, which can produce a Parkinsonian syndrome in man (Davis *et al.*, 1979). *In vitro* studies have shown that MPP^+ (1-methyl-4-phenylpyridinium ion) and its precursor MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) can competitively inhibit CYP2D6 and are possible substrates (Fonne-Pfister *et al.*, 1987).

Another disease related gene is *CYP11B1* encoding an 11ß-hydroxylase. It has been found that inherited deficiency of 11ß-hydroxylase can cause congenital adrenal hyperplasia, a disorder of cortisol biosynthesis (New *et al.*, 1989). The missense mutations reported to be associated with 11ß-hydroxylase deficiency are four base pair mutations and a base pair deletion in exon 2 of the *CYP11B1* gene. The deletion causes a frameshift and premature termination of the protein. Each of these point mutations has been shown by *in vitro* transfection to abolish 11ß-hydroxylase activity (Curnow *et al.*, 1993).

1.5 Aims of the Study

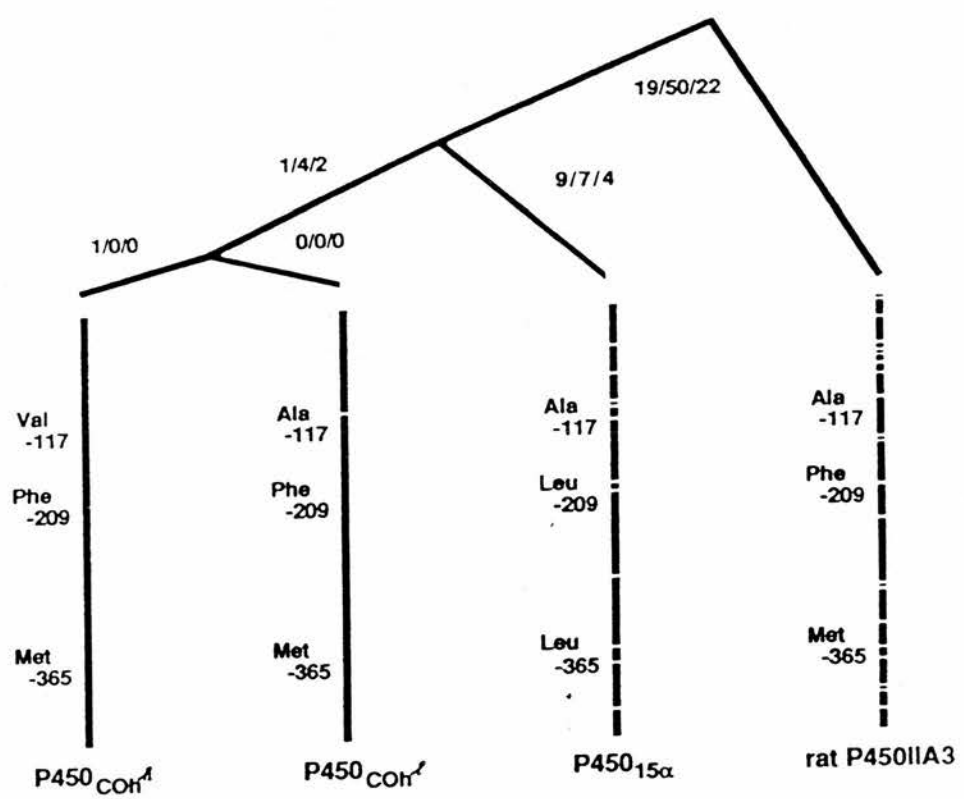
Our present knowledge of the P450 system already indicates that many factors determine P450 levels in an individual, including genetic background, dietary habits, hormonal level and exposure to foreign compounds that act as inducers or repressors. The great inter-individual differences in activity of P450 can influence individual susceptibility to clinical drug treatment, chemically induced disease and cancers. If we can decide whether the level of a particular P450 in an individual is at a dangerously high or low level, or the individual differences are associated with cancer susceptibility and environmentally linked disease, it will be possible either to adjust the drug exposure or to avoid the exposure to harmful substances. For instance, it may be possible to predict whether a tobacco smoker is more or less susceptible to lung cancer.

All of these approaches will be on the basis of a thorough understanding of the human P450 system.

In addition, one of the most critical concerns of pharmacology is to define the dose-response relationships of a drug. A large amount of information has clearly demonstrated that drug metabolism and the regulation of drug-metabolising enzyme activity are often different in man compared to experimental animals. Humans appear to have their own unique set of P450 genes. The inability to correlate xenobiotic metabolism in humans and in rodents stresses the importance of a thorough analysis of human P450s. Therefore, human-like systems to study drug metabolism must be developed.

The CYP2A subfamily is one of the important gene families involved in the metabolism of xenobiotics, and has been extensively studied in rodents. In rat, three cDNAs have been isolated and sequenced (Nebert *et al.*, 1991). The enzymatic activity of CYP2A1 was found to specifically hydroxylate testosterone at the 7α position. Although *CYP2A2* is highly homologous to CYP2A1 by deduced amino acid sequence, it does not have the same degree of specificity to testosterone hydroxylation. The expression levels of both genes are also regulated differently during development and after treatment with the carcinogen 3'-methylcholanthrene. Rat *CYP2A3*, appears to be expressed only in lung, and the expression level is induced by 3'-methylcholanthrene (Kimura-S. *et al.*, 1989). In mice, two cDNAs, *Cyp2a-4* and *Cyp2a-5,* have been isolated (Burkhart *et al.*, 1990). The sequences of both cDNAs exhibit 98% similarity, however, the encoded enzyme activities depend on the three amino acids at positions 117, 209, and 365 (Lindberg *et al.*, 1989). Interestingly, only one amino acid change results in a high or low activity of coumarin 7-hydroxylase in different mouse strains (Lindberg *et al.*, 1992). This result indicates that during evolution, amino acid substitutions have selectively occurred at positions which alter substrate specificity of enzyme and increase enzymatic activity (Figure 1.8).

Figure 1.8 Phylogeny of the CYP2A family. At the top is the most parsimonious phylogeny. The numbers next to each branch are the inferred minimum numbers of nonsynonymous/synonymous/noncoding substitutions that occurred in each branch. The vertical bars depict the conceptual translations of the cDNAs. Sites that differ from P450coh^{H} are marked by open ticks. The amino acids at the three sites shown to effect coumarin 7-hydroxylase and steroid 15α-hydroxylase activities are also indicated. P450Coh^{h}, high coumarin 7-hydroxylase activity; P450Coh^{l}, low coumarin 7-hydroxylase activity and $P450_{15\alpha}$, steroid 15α-hydroxylase (adapted from Gonzalez, 1992; Lindberg and Negishi, 1992).



Two distinct *CYP2A* cDNAs have been isolated from a human cDNA library (Yamano *et al.*, 1990), one of which (*CYP2A6*) has been identified as encoding a coumarin 7-hydroxylase (Miles *et al.*, 1990). The second one, designated *CYP2A7*, encodes a protein containing a complete reading frame and exhibits 96% nucleotide sequence similarity with *CYP2A6*. However, the catalytic activity of it is unknown.

It has been found that human liver samples exhibit a great inter-individual variability in levels of CYP2A mRNAs, protein and coumarin 7-hydroxylase activity (Yamano *et al.*, 1990; Miles *et al.*, 1990; Yun *et al.*, 1991). Up to 144-fold variability of CYP2A protein(s) was detected in human liver samples (Maurice *et al.*, 1991). A possible reason is that *CYP2A* genes are polymorphically expressed in humans, and the mutant alleles in the population result in the marked inter-individual variability in the activity of coumarin 7-hydroxylase. If polymorphisms of *CYP2A* genes exist indeed, it will be important to determine whether the expressions of *CYP2A* genes are associated with an increased risk for chemically induced cancer because the enzyme is involved in the metabolic activity of aflatoxin B_1 and N-nitrosodiethylamine (NDEA) (Crespi et al., 1990; Yamazaki *et al.*, 1992; Tiano et al., 1993)

However, the human *CYP2A* subfamily has not been extensively investigated, and no human *CYP2A* gene has ever been isolated. In order to characterise the genetic and environmental factors involved in the inter-individual variability in the expression levels of human *CYP2A* genes, and to study their transcriptional regulation, the *CYP2A* genes need to be isolated and their structure and promoter element(s) need to be thoroughly characterised. On the basis of studies relating to human P450 subfamilies 2A and 2B in this laboratory (Miles *et al.*, 1989; 1990), the aims of this thesis were therefore as follows:

1). To isolate human cytochrome P450 genes - in particular, to isolate the genes in the *CYP2A* family.

2). To characterise the gene structure, including intron-exon junctions and promoter region in order to understand the regulation of the *CYP2A* gene.

3). To define the factors involved in the transcriptional regulation of the *CYP2A* gene and possible induction by xenobiotics.

4). To study the molecular mechanisms of inter-individual variation in the expression of the *CYP2A* gene and the catalytic activity of the enzyme encoded by the gene.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 Commercial material sources

Most chemicals were supplied by BDH limited, Burnfield Avenue, Thornliebank, Glasgow, G46 7TP and Sigma Chemical Company Limited, Fancy Road, Poole, Dorset, BH17 7NH. A list of chemicals including tissue culture materials and suppliers is given in Appendix 1.

Human genomic library was supplied by CLONTECH Laboratories, Inc.

2.1.2 Non-commercial materials

Cytochrome P4502A antibody is a polyclonal antiserum to rat P450IIA1, supplied by Prof. C.R. Wolf. The antiserum has been extensively characterised in this and other laboratories, and cross-reacting protein was quantified for comparative purpose by scanning the autoradiographs. Standard curves establishing the linearity of the method were made by comparing the band intensities obtained on serial dilution of a microsomal sample (Wolf *et al.*, 1988; Miles *et al.*, 1990; Forrester *et al.*, 1992).

Cosmid library was a gift from Dr. D. Kioussis (Kioussis *et al.*, 1987). It is a human acute lymphocytic leukaemia genomic DNA library in an EBV-based cosmid vector cos202.

2.2 Cell Culture

2.2.1 Cell lines

HepG2 cells were grown in Dulbecco's minimal essential medium supplemented with 15% (v/v) fetal calf serum and antibiotics (penicillin 50 IU/streptomycin 50 μg/ml). The HepG2 cell line was derived from a childhood hepatoblastoma and has been shown to

retain many of the different features of human liver including cytochrome P450-dependent monooxygenase activity (Knowles *et al.*, 1980).

COS-7 is a simian virus 40-transformed monkey kidney fibroblast cell line. The cells were grown in Dulbecco's minimal essential medium supplemented with 10% (v/v) fetal calf serum and antibiotics (penicillin 50 IU/streptomycin 50 µg/ml).

Human skin fibroblast cell line (FEK4) was kindly donated by Dr. S. Keyse. The cells were cultured in DMEM medium supplemented with 10% (v/v) fetal calf serum and antibiotics (penicillin 50 IU/streptomycin 50 µg/ml).

2.2.2 Conditions for cell culture

The cell culture methods used were as described by Freshney (1987). All culture work was conducted in a Class 2 Biological Safety Cabinet, MDH Ltd., Walworth Road, Andover, Hampshire. The cabinet was U.V. sterilised when not in use and cleaned at regular intervals.

2.2.3 Freezing cells for storage and retrieving stocks

Upon reaching about 80% of confluence, cells were harvested using a solution containing 0.125% (w/v) trypsin and 0.01% (w/v) EDTA until the monolayer detached. The trypsin was then quickly diluted out by adding growth medium containing 10% (v/v) fetal calf serum (specific for each particular cell line) and the cells were spun down for 5 min at 1500 rpm in a MSE Microcentaur centrifuge. The cell pellet was then resuspended at a density of 0.5-1.0 x 10^6 cells/ml in 90% (v/v) fetal calf serum, 10% (v/v) DMSO. Aliquots of 1 ml were frozen at -70^{0}C overnight and then transferred to liquid nitrogen for long-term storage. DMSO permeates cells rapidly and maintains long-term viability of cell lines. To retrieve cells from storage, an aliquot was thawed at 37^{0}C and carefully resuspended in the appropriate growth medium pre-warmed to 37^{0}C. The cells were then seeded into a 80 cm^2 tissue-culture flask. The cells were

allowed to adhere overnight and re-fed the next morning to eliminate any dead cells or other debris from the flask.

2.2.4 Feeding cells

Cells were fed with specific media as required by each cell line. Most cell lines required re-feeding every 2 or 3 days to maintain optimal pH and essential growth requirements.

2.2.5 Sub-culture of cells

Just before reaching confluence, cells were sub-cultured by washing them twice in PBS and harvesting them using a solution containing 0.125% (w/v) trypsin and 0.01% (w/v) EDTA until the cells detached. Fresh medium with serum was then added to inhibit the trypsin and the cell suspension was spun at 1500 rpm for 5 min. The cell pellet was resuspended in fresh medium. To ensure accurate seeding, when required, the cells were counted using a Neubauer haemocytometer and seeded into fresh flasks. Most parental cell lines were sub-cultured every 3-4 days at a dilution of 1:10 to 1:15.

2.2.6 Sterility

Culturing of cells was carried out in a sterile cabinet and aseptic techniques were employed at all times. Solutions were either sterilised in a Laboratory Thermal Equipment 225 EH autoclave or filter-sterilised (0.2 μm pore size) before use. Testing the sterility of media and other solutions was routinely done by inoculation of the test solution into sterile L-broth [2%, (w/v) Bactotryptone, 1% (w/v) Bactoyeast extract, 2% (w/v) NaCl] and incubating for 4-7 days at 37^0C. A turbid L-broth test was an indication of some form of contamination and the solution could be discarded before use. Regular mycoplasma tests were done by Mr. W. Christie, MRC Clinical and Population Cytogenetics Unit, Western General Hospital, Crewe Road, Edinburgh. The mycoplasma test involved staining with Hoescht 33258 fluorescent dye (Chen, 1987) and growth on selective broth (Taylor and Robinson,1978).

2.2.7 Preparation of fraction for cytochrome P450 estimation

Cells were grown in 80 cm^2 flasks, seeded at 3 x 10^6 cells per flask. Following treatment as before, the cells were washed twice in ice-cold PBS and harvested using a solution containing 0.125% (v/v) trypsin, 0.01% (v/v) EDTA. Cold PBS was added to the flasks following detachment of the cells and the cell suspension spun at 1500 rpm for 5 min. The cell pellet was then resuspended in 0.5 ml cold, filter-sterilised phosphate buffer, pH 7.4 (10 mM Na_2HPO_4, 2 mM $MgCl_2$, 2 mM DTT, 1 mM EDTA). The cells were sonicated using an MSE Soniprep (amplitude 12 μm, 5 sec, twice) with samples kept on ice. Cell disruption was assessed visually by checking an aliquot under the microscope at 10 x magnification. The disrupted suspension was then spun at 6000 rpm for 5 min in a microfuge. The supernatant was retained whilst the pellet was re-suspended in 200 μl of the phosphate buffer. Both fractions were assayed separately.

2.3 Analysis and Enzymatic Manipulation of DNA

2.3.1 Materials and solutions

Electrophoresis-grade agarose, and 1 kb DNA molecular weight markers (GIBCOBRL Ltd).

10 x loading buffer: 20% (w/v) Ficoll 400, 0.1M Na_2EDTA, pH 8.0, 1.0% (w/v) sodium dodecyl sulphate, 0.25% (w/v) Bromphenol Blue and 0.25% (w/v) Xylene.

TE buffer: 10 mM Tris-Cl (pH8.0), 1 mM EDTA.

5 x TBE buffer: 54 g Tris base, 27.5 g boric acid and 20 ml 0.5 M EDTA (pH 8.0).

20 x TAE buffer: 96.8 g Tris base, 22.8 ml glacial acetic acid and 40 ml 0.5 M EDTA (pH 8.0).

Electrophoresis working solution: 0.5 x TBE or 1 x TAE with 0.3 μg/ml ethidium bromide.

20 x SSC: Dissolve 175.3 g of NaCl and 88.2 g sodium citrate in 800 ml of H_20. Adjust the pH to 7.0 with a few drops of a 10 N solution of NaOH, and then add water to 1 litre.

10 x CIP buffer: 0.2 M Tris-Cl, pH 8.0, 10 mM $MgCl_2$, 10 mM $ZnCl_2$.

OLB (oligo-labelling buffer) consists of a mixture of solutions A, B and C mixed in a ratio of 2:5:1 and was routinely stored at -20^{0}C.

Solution A

2 M Tris-HCl, pH 8.0	625 μl
water	82 μl
2-mercaptoethanol (99% min)	18 μl
$MgCl_2$ (1 M)	125 μl
10 mM dATP, dGTP, dTTP	50 μl (of each)

Solution B: 2 M Hepes, adjust pH to 6.6 with NaOH and stored at 4^{0}C.

Solution C: Hexadeoxyribonucleotides suspended in 3 mM Tris-Cl 0.2 mM EDTA, pH 7.0 at a concentration of 90 A units/ml and stored at -20^{0}C.

Denhardt's reagent: 0.02% (w/v) polyvinylpyrrolidine, 0.02% (w/v) bovine serum albumin (BSA) and 0.02% (w/v) Ficoll 400.

2.3.2 Quantification of DNA by spectrophotometric measurement

DNA concentration was determined by measuring the UV absorption at 260 nm. For double-stranded DNA, 5.0 μl sample was diluted to 1 ml with deionized water. For single strand oligonucleotides, 2.0 μl sample was diluted to 1 ml and mixed thoroughly. The absorption at 260 nm was read and the concentration was calculated according to the following equation:

DNA (μg) = A_{260} x 10

where A_{260} is the absorption at 260 nm. The above equation is based on the assumption that 1 A_{260} corresponds to approximately 50 μg/ml for double-stranded DNA or 20 μg/ml for single-stranded oligonucleotides. To assess the purity of the sample, the absorption at 280 nm is also obtained and the ratio between the readings at A_{260} and A_{280} provides an estimate of the purity of the DNA. A pure preparation of DNA sample has an A_{260}/A_{280} value of 1.8.

2.3.3 Agarose gel electrophoresis

The desired amount of agarose (0.8-2.0%, w/v) was added to a volume of electrophoresis buffer and the agarose was melted in a microwave oven. The melted gel was cooled to 50^{0}C before pouring. DNA samples were prepared by adding 1/10 volume of 10 x loading buffer and loaded into the wells with a pipette. 1 kb DNA molecular weight standards were used and gel electrophoresis carried out at 2 to 8 V/cm. Separation was monitored by the migration of the dyes, and the DNA visualised using UV light and photographed.

2.3.4 Isolation and purification of DNA restriction fragment from low melting agarose gel

DNA was digested with appropriate restriction enzyme(s), digested DNA fragments were loaded onto a 1% to 2% (w/v) low melting gel (dependent on the size of DNA fragment), and electrophoresed in 1 x TAE buffer (see Section 2.3.3). After electrophoresis, the target band was carefully cut out and purified using a GlassMAX DNA Isolation Spin Cartridge System (GIBCO BRL Ltd). In detail, for 0.1 g recovered gel 450 μl of binding solution (6M sodium iodide, NaI) was added, and the mixture was heated at 65^{0}C until the agarose gel was fully dissolved. The 550 μl of DNA/NaI mixture was added to the spin cartridge, centrifuged at 13,000g for 20 seconds. The spin cartridge was washed by 0.4 ml washing solution for three times,

then moved to a fresh tube. 40 µl of the TE buffer preheated to 65^{0}C was added into the spin cartridge, and then the spin cartridge with the tube was centrifuged at 13,000 rpm for 20 seconds to elute the DNA. The purified DNA was used in subsequent manipulations.

2.3.5 Subcloning of DNA fragment

Ligation: vector and insert DNA (5 to 10 µg) was digested with appropriate restriction enzyme(s), then the target DNA fragment was isolated and purified by low melting agarose gel (section 2.3.4). For dephosphorylation of 5' phosphate 1/10 volume of 10 x CIP buffer and 1U CIP (Promega) was added and incubated 60 min at 37^{0}C. For blunt end conversion 1/10 volume of a solution containing all 4 dNTPs (0.1 mM) and 2-5 U Klenow large fragment of E.coli DNA polymerase was added and incubated 30-60 min at 22^{0}C. After the reaction was complete, the DNA fragment was purified by GlassMAX DNA Isolation Spin Cartridge System (section 2.3.4). Ligation reactions were set up as follows and incubated at 15^{0}C water bath overnight. The ligated recombinant plasmid DNA was transformed into competent cells according to the protocol in section 2.4.5.

Vector DNA	100 ng
Insert DNA	X ng
T_4 DNA ligase	1u (Weiss units)
10 x ligation buffer*	1.0 µl
H_2O	to final volume 10 µl

X: the calculated amount for the molar ratio of vector to insert, 1:1 to 1:3.

*10x ligation buffer: 500 mM Tris-HCl (Ph 7.6), 100 mM $MgCl_2$, 10 mM ATP, 10 mM DTT.

2.3.6 Southern blotting

The amount of DNA in a Southern blotting (Southern, 1975) depends both on the complexity of the DNA and the probe to which it will be hybridised. An amount of 1 ng or less of a plasmid DNA will be sufficient to yield a signal that can be detected within hours. However, 10 to 20 μg of mammalian total genomic DNA is usually required to yield a signal that can detected within 1 to 2 days.

DNA (10 ng to 20 μg) was digested completely, loaded onto an agarose gel and electrophoresed for 4 to 12 hr. Following electrophoresis the gel was photographed, transferred to a tray containing 500 ml denaturation solution (1.5 M NaCl, 0.5 M NaOH) with constant, gentle agitation for 60 min. The gel was rinsed twice in water after pouring off the denaturation solution, and then 500 ml neutralisation solution (0.5 M Tris-Cl, pH8.0, 1.5 M NaCl) was added and the tray was rocked again for 60 min. Transfer to Amersham Hybond N membrane was achieved by capillary action in 10 x SSC for 16-20 h, and the DNA was fixed to the nylon membrane by exposing the filter to ultra-violet irradiation (254 nm) for 30 seconds using a Stratalinker 2400. This method is time saving and enhances the hybridisation signal compared to conventional oven-baking (Khandjian, 1987).

2.3.7 Preparation of radiolabelled probes and hybridisation

The DNA or cDNA fragment used as a probe was isolated and purified according to the protocols in Section 2.3.4. A small amount of a DNA fragment (50-100 ng) was labelled to high specificity using the method of Feinberg and Vogelstein (1983). This method was used to generate probes from denatured double-stranded DNA. The purified DNA, mixed with a molar excess of random primers, was denatured and synthesis was carried out using the Klenow polymerase I. This enzyme lacks 5'-3' exonuclease activity so that the product is synthesised exclusively by primer extension.

DNA (0.1 µg) in a volume of 10 µl was denatured by boiling for 5 min and then 3 µl OLB buffer was added. The reaction was kept at room temperature for 10 min, then 3 µl of BSA (Fraction 5, Sigma, 10 mg/ml), 4 µl H_2O, 2 µl Klenow polymerase I (4 units), and 3 µl (α-^{32}P)-dCTP (sp. act. >3000 Ci/mmol) were added and mixed. The radiolabelling reaction was allowed to proceed for 5 h to 16 h. Following incorporation of radiolabel, the DNA was diluted to 100 µl with TE buffer containing 15 mM EDTA and denatured for 5 min at 100^{0}C. 1/3 volume of reaction mixture was added to the hybridisation solution as quickly as possible. To estimate the percentage of radioactivity incorporated into the DNA, a 2 µl aliquot was removed from the diluted reaction mixture and spotted onto DE 81 Whatman filter paper. A chromatography experiment was run in 0.3 M ammonium formate, pH 8.0 for 20-60 min. The filter paper was wrapped in Saran parafilm and exposed to Kodak X-Omat AR-5 film for 10 to 60 min. Radioactivity incorporated into the DNA appeared as a spot at the origin whereas unincorporated nucleotides was eluted up the paper with the solvent front. Incorporation is routinely between 50-70%.

The membrane with DNA was placed in a glass tube with prehybridisation mixture (5 x SSC, 5 x Denhardt's solution, 0.1% (w/v) SDS, 150-200 µg denatured salmon sperm DNA) for 2-4 h at 65^{0}C. Denhardt's reagent was used to block non-specific attachment of the probe to the membrane surface. Following prehybridisation, denatured probe was added directly to the prehybridisation solution and allowed to hybridise to the DNA for 16-18 h at 65^{0}C. After hybridisation, the membrane was washed several times in 0.1 x SSC, 0.1% (w/v) SDS at room temperature, then at 65^{0}C for 30-60 min to eliminate cross-reactivity to similar DNA sequences. The filter was wrapped in Saran wrap and exposed to Kodak X-Omat AR-5 film at -70^{0}C with intensifying screens.

2.4 Culture of Escherichia coli and Preparation of Plasmid DNA

Initially the procedure used for large-scale and small-scale preparation of plasmid DNA was a modification of the methods of Birmboin and Doly, (1979), Ish-Horowicz and Burke, (1981). Qiagen kits were later employed.

2.4.1 Materials and solutions

L-broth medium (LB):

Difco bactotryptone 10 g/l

Difco yeast extract 5 g/l

NaCl 5 g/l

L-broth-agar: As for L-broth but including 12 g/l Difco agar.

Solution I: 50 mM glucose, 25 mM Tris-Cl (pH8.0), 10 mM EDTA (pH8.0).

Solution II: 0.2 N NaOH (freshly diluted from a 10 N stock) and 1% (w/v) SDS.

Solution III:

5 M potassium acetate	60 ml
glacial acetic acid	11.5 ml
H_2O	28.5 ml

The resulting solution is 3 M with respect to potassium and 5M with respect to acetate.

CsCl/ethidium bromide solution (Density is between 1.55 and 1.59 g/ml):

TE buffer	100 ml
CsCl	110 g
ethidium bromide (10 mg/ml)	10 ml

Buffer 1 (for competent cells preparation): 30 mM Potassium Acetate, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$ and 15% (v/v) glycerol. This solution should be prepared fresh and filter-sterilised for each use.

Buffer 2 (for competent cells preparation): 10 mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$ and 15% (v/v) glycerol.

2.4.2 Media used for bacterial culture

L-broth and L-agar were used for the culture of DH5α in liquid and solid phase respectively. L-agar plates with streaked out colonies were stored at 4^0C for up to 4 weeks.

2.4.3 Storage of bacteria

Bacterial cells were grown in L-broth overnight with appropriate antibiotic selection. Glycerol was added in the overnight culture to a final concentration of 15% (v/v) and 1 ml aliquots were frozen at -70^0C.

2.4.4 Preparation of competent cells

The following procedure is a variation of the Hanahan method (1983), and all culture media contain the appropriate antibiotic. A single colony from an L-agar plate was inoculated into 10 ml LB medium and cultured overnight at 37^0C with vigorous shaking. Inoculate 200 ml of LB medium with 2 ml of overnight cultured cells in a 1.0 litre flask, shake the culture at 37^0C until the OD_{600} reaches 0.35, chill the cells in ice water for 1 h and collect the cells by centrifugation at 2,500 rpm for 12 min at 0^0C. Resuspend the cells in 100 ml of ice-cold buffer 1, then keep the cells on ice for 15 min, centrifuge at 2,500 rpm at 0^0C for 12 min and gently resuspend the pellet in 20 ml of ice cold buffer 2. Chill the resuspended solution on ice for 15 min. Aliquot the cells in 0.2-1.0 ml quantities, freeze in dry ice and store at -80^0C.

2.4.5 Transformation of competent cells

An aliquot (0.2 ml) of competent cell was thawed on ice. Ligated plasmid DNA (10 to 200 ng) was added, and the cells was incubated on ice for 30 min, and heat shocked at 42^{0}C water bath for 1.5 min. For some strains, this treatment has been reported to increase transformation efficiency. Following the heat shock, cool the tube on ice for 1 minute, add 0.8 ml of LB medium and shake at 37^{0}C for 1 hr to allow the cells to recover. The cells could be plated directly or concentrated (centrifuge and resuspend the cells in 100-200 µl of LB medium) before plating on an LB plate containing appropriate antibiotic and incubated at 37^{0}C for 12-16 h.

2.4.6 Preparation of plasmid DNA

(A) Large-scale preparation of plasmid DNA or cosmid DNA

A single colony from an L-agar plate was inoculated into 10 ml of L-broth plus appropriate antibiotic and incubated overnight at 37^{0}C with shaking. The culture was seeded into 200 ml of pre-warmed L-broth plus appropriate antibiotic in a 1 litre conical flask and incubated overnight at 37^{0}C with vigorous shaking. The cell suspension was spun at 5,000 rpm at 4^{0}C for 10 min in a Sorvall RC 5B rotor. The bacterial pellet was then resuspended in 20 ml of solution I with 2 mg/ml lysozyme, and kept on ice for 30 min to lyse the bacterial cell wall. Solution II was then added, mixed and left on ice for 5 min to further disrupt the bacterial cell membrane. To remove the SDS and the proteins, solution III was added and the mixture was left on ice for a further 20 min. The white precipitate was pelleted by centrifugation at 10,000 g, 4^{0}C, for 15 min. The supernatant, containing the plasmid DNA, was retained and isopropanol (0.7 volumes) was added. The DNA was precipitated by keeping the solution for 30 min at room temperature followed by a 10 min spin (10,000 g, 4^{0}C).

The DNA pellet was resuspended in 15.5 ml of TE (10 mM Tris-Cl, pH 8.0, 1 mM EDTA), to which caesium chloride (14.4 g) and ethidium bromide (10 mg/ml, 2.3 ml)

were added. After the suspension was centrifuged (200,000g, 16 hour, 15^{0}C), the DNA bands were visualised using a short-wave transilluminator. Using a hypodermic needle inserted into the side of the ultracentrifuge tube, the band of circular plasmid DNA was collected as described by Maniatis (1989). For transfection, the plasmid DNA was centrifuged a second time. The ethidium bromide was extracted using water-saturated butanol and the DNA solution dialysed against three changes of 1,000 volumes of TE at 4^{0}C for 16 h with stirring. The DNA was precipitated at -70^{0}C for 60 min by adding 2 volumes cold ethanol, 0.1 volume 3 M sodium acetate, pH 4.8 and spinning at 10,000g, 4^{0}C for 30 min. The DNA pellet was finally washed in 70% (v/v) ethanol, dried briefly by desiccation and redissolved in a suitable volume of TE buffer and kept at 4^{0}C.

(B) Plasmid DNA preparation by QIAGENE column method

The initial steps of the protocol were essentially the same as for part A. Following addition of solution III to precipitate SDS, the mixture was centrifuged at 4^{0}C for 30 min, 20,000g. The supernatant was carefully but promptly removed to obtain a particle-free clear lysate. A QIAGENE column was equilibrated with 10 ml of buffer QBT (750 mM NaCl, 50 mM MOPS, 15% (v/v) ethanol, pH 7.0, 0.15% (v/v) Triton X-100) and allowed to empty by gravity flow. The supernatant containing plasmid DNA was then applied to the column and allowed to enter the resin by gravity flow. The column was then washed with 2 x 30 ml buffer QC (1.0 M NaCl, 50 mM MOPS, 15% (v/v) ethanol, pH 7.0). At pH 7.0, the solubility of DNA is reduced and it will bind to the resin. The DNA was eluted with 15 ml of buffer QF (1.25 M NaCl, 50 mM MOPS, 15% (v/v) ethanol, pH 8.2). The DNA was then precipitated with iso-propanol (0.7 volumes) and centrifuged at 4^{0}C for 30 min, 10,000g. The DNA pellet was finally washed in 70% (v/v) ethanol, dried briefly by desiccation and redissolved in a suitable volume of buffer and kept at 4^{0}C.

(C) Small-scale preparation of plasmid DNA

1.5 ml of overnight culture was poured into a microfuge tube, and centrifuged at 12,000g for 2 min at room temperature. The supernatant was removed by aspiration, and the pellet was resuspended in 100 μl ice-cold Solution I by vigorous vortexing. 200 μl solution II was added, mixed by inverting the tube several times, then 150 μl of solution III was added to the tube and the contents were quickly mixed by inverting the tube several times. 450 μl phenol:chloroform (1:1) solution was added, mixed by vortexing and centrifuged at 13,000 rpm for 2 min. The supernatant was transferred to a fresh tube. Then 2 volumes of ethanol at room temperature were added and mixed to precipitate the plasmid DNA for 2 min. After centrifuging at 13,000 rpm for 10 min, the supernatant was removed and the pellet was washed with 1 ml of 70% ethanol. The pellet was allowed to dry at room temperature, and redissolved in 50 μl of TE buffer containing DNase-free pancreatic RNase (50 μg/ml). The plasmid DNA prepared in this method can be used in the restriction enzyme analysis.

2.5 Screening of Human Genomic Library

2.5.1 Human genomic cosmid library

A human genomic DNA library was constructed by Dr. Kioussis *et al.* (1987). Figure 2.1 shows the structure of the EBV-based shuttle cosmid vector cos202. The human genomic DNA fragments (25-40 kb) were inserted into the *Bgl*II site.

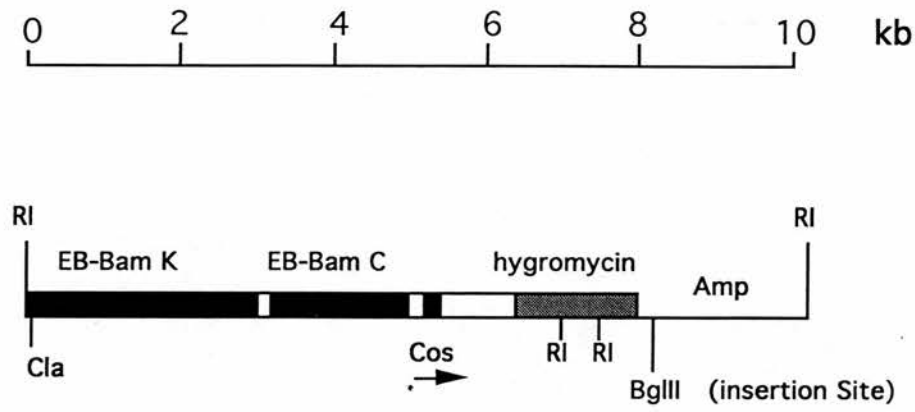


Figure 2.1 Structure of cosmid vector cos202. The EB-Bam K and EB-Bam C are fragments from Epstein-Barr virus (EBV). Amp, the ampicillin resistance gene, and RI (*EcoR*I), Cla (*Cla* I) and BglII, the restriction enzyme sites.

2.5.2 Screening the library for CYP2A genes

(A) Screening of cosmid library:

The cosmid library was plated onto Hybond-N filters at a density of 100,000 colonies per plate as follows: a dry Hybond filter was layered on to a surface of freshly made LB-ampicillin plate to wet it. Cells were spread over the filter in a volume of 0.5 ml per plate using a bent glass rod and then incubated at 37^0C until colonies were just visible and then stored in a refrigerator.

To make replicas, the master filter was lifted from its plate and placed colony side up on several layers of Whatman filter paper on a flat rigid surface. A new piece of Hybond filter was placed on a freshly made LB-ampicillin plate to allow it to become damp, then lifted from the plate and placed on the top of the master filter and pressed hard over the surface of the filter using the palm of the hand. A series of key holes were made through the filters in an irregular pattern using a syringe needle dipped in water-proof ink. The filter was peeled apart, replaced on its plate and incubated at 37^0C. The master plate was incubated at 37^0C until the colonies reappeared, and then the next replica was

made. Duplicate replicas were used for screening. For screening, the replica was laid on a piece of 3 MM paper soaked in 0.5 M NaOH for 4 min, transferred to a similar piece of filter paper soaked in 1 M Tris-Cl pH 7.6, left for 4 min, and then transferred to a further piece of filter paper soaked in 1.5 M NaCl, 0.5 M Tris pH 7.6, for 4 min. Following vigorous removal of bacterial debris from one side to the other by using tissue paper soaked in 2 x SSC, 0.1% (w/v) SDS, the replica was washed briefly in 2 x SSC, air dried and DNA was fixed by U.V.

(B) Screening of bacteriophage library

Place 50 ml LB supplemented with 0.2% (w/v) maltose in a sterile 250 ml flask and inoculate with a single bacteria colony. Grow the culture overnight at 37^0C with shaking. Pellet the cells by centrifugation at 4000 rpm for 10 min at room temperature and resuspend the cells in 20 ml of 10 mM $MgSO_4$. Store the cell suspension at 4^0C.

Mix an aliquot of the recombinant phage and an aliquot of the plating bacteria (see Table 2.1) in a culture tube and incubate 20 min at 37^0C. Add molten (47^0C) top agarose (0.7%) to the tube and pour the mixture to a LB plate. Disperse bacteria and agarose on the plate by tilting the plate back and forth. Incubate the plate at 37^0C until plaques cover the plate but were not confluent, chill the plate at 4^0C for at least 1 h before applying filter. Remove the plate from refrigerator and place a nylon filter neatly onto the surface of the top agarose. This was accomplished by touching first one edge of the filter to the agarose and progressively laying down more of the filter as it wets. Bubbles should be avoided. Mark the filter in three or more asymmetric locations by stabbing through it and into the agar beneath with an 18-gauge needle attached to a syringe containing waterproof black drawing ink. After 30-60 seconds, remove the filter carefully from the plate with blunt, flat forceps. Place the filter, DNA side up, on a Whatman 3MM paper saturated with 0.5 M NaOH, 1.5 M NaCl for 5 min, then transfer the filter, DNA side up, to another 3MM paper saturated with 0.5 M Tris (8.0),

1.5 M NaCl for 5 min. Rinse the filter in 2 x SSC, and place it, DNA side up, on paper towers to dry for 30 min. Up to 5 replicas can be made from each plate.

The human genomic DNA libraries were screened using a 0.7 kb *CYP2A6* cDNA (Miles *et al.*, 1988) as a probe. The hybridisation and washing conditions were as described for Southern blotting in section 2.3.6.

Table 2.1 Recommended mixtures for plating bacteriophage libraries (Adapted from Ausubel *et al.*, 1987)

	Plate Size		
LB plate ingredient	82 mm	150 mm	245 x 245 mm
Plating Bacteria (ml)	0.2	0.5	2.0
Phage, pfu	5000	20,000-30,000	150,000
Top agarose (ml)	3.0	7.0	30

2.6 DNA Sequencing

The dideoxy chain-termination method (Sanger *et al.*, 1977) was used with (α-^{35}S)-dATP (400 Ci mmol^{-1}) to sequence double-strand DNA cloned in pUC18, pCAT-basic plasmid (Promega) or other vectors. Sequences were compiled and analysed using Gene Jockey software. The sequencing procedure was as follows.

Plasmid DNA was made from 10 ml overnight culture according to the small-scale preparation of plasmid DNA in section 2.4.4.C. Following ethanol precipitation, redissolve the DNA pellet in 150 µl TE buffer, transfer an aliquot of 50 µl to a fresh tube, then add RNase A to 100 µg/ml. Keep the tube at 37^{0}C overnight, then add 30 µl of PEG solution [20% (w/v) polyethylene glycol 6000/ 2.5 M NaCl], mix and chill the tube on ice for at least 1h to precipitate DNA. The aim of PEG precipitation is to

remove the digested RNA from the DNA pellet. After centrifugation, wash the DNA pellet with 70% (v/v) ethanol at room temperature and redissolve the DNA in 54 µl TE buffer (for three sequence reactions). Transfer an aliquot of 18 µl to a fresh tube, add 2 µl of 2 N NaOH to denature the DNA at room temperature for 5 min, and then add 8 µl of 5M ammonium acetate (pH 7.5) to neutralise the reaction. Add 4 volumes of ethanol at 0^0C to the tube, mix, and store the solution at -80^0C for 30 min. Recover the denatured plasmid DNA by centrifugation at 13000 rpm for 12 min in a microfuge. Carefully remove and discard the ethanol. Leave the tube open on the bench until the last traces of ethanol have evaporated. Dissolve the DNA in 7 µl of water. The sequencing reaction is carried out using a Sequenase Version 2.0 DNA Sequencing Kit and according to the protocols supplied by United States Biochemical.

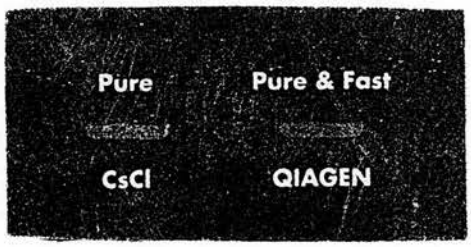
Briefly, add 2 µl reaction buffer, 1 µl primer, anneal by heating 2 min at 65^0C and then cool slowly to below 35^0C over 30 min. While cooling, dilute the Labelling Mix 10-fold with distilled water (e.g. 1 µl of Mix combined with 9 µl of water). Dilute the sequenase 2.0 enzyme 1:8 in ice-cold enzyme dilution buffer. To the annealed template-primer (10 µl) mixture, add diluted Labelling solution 2 µl, DTT 1 µl, (α-^{35}S)dATP (>400Ci/mmol) 0.5 to 1 µl and diluted sequenase 2 µl (3 units). Mix the reaction thoroughly and incubate for 2-5 min at room temperature. Remove 3.5 µl reaction mixture to the tube labelled G, in which 2.5 µl ddGTP has been added and prewarmed at 37^0C. Similarly transfer 3.5 ml reaction mixture to A, T and C tubes with 2.5 µl ddATP, ddTTP and ddCTP, respectively. Mix and return them to the 37^0C water bath. After incubation for 2-5 min, add 4 µl of stop solution to each of the termination reactions and mix. When the 6% denaturing gel was ready for loading, heat the samples to 75-85^0C for 2 min and load immediately on the gel with 2-4 µl in each lane.

2.7 Strategies for Studying Gene Regulation and Expression

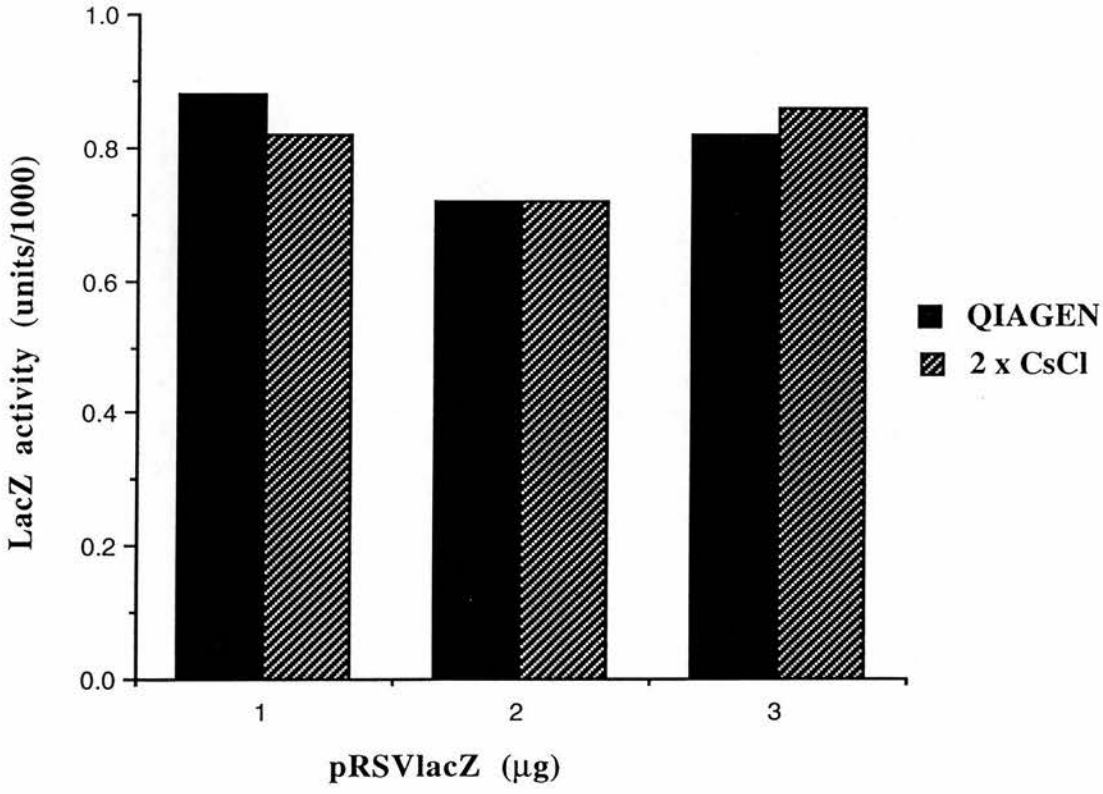
The DNAs for transfection were prepared by the alkaline lysis method and purified either by filtration (QIAGENE), or CsCl density gradient centrifugation (Section 2.4.6,

Figure 2.2 Comparison of plasmid DNA prepared by caesium chloride/ultracentrifugation or by QIAGEN filtration.

A: Electrophoresis of DNA prepared with different methods.



B: Comparison of transfection efficiency. NIH3T3 cells were transfected with the plasmid pRSVlacZ which were prepared with QIAGEN or by CsCl. Data from Ehlert *et al.*, (1993) *Bio Techniques*, **14**, 546.



A and B). The transfection efficiencies of DNAs prepared by different methods is shown in Figure 2.2.

2.7.1 Buffers

All buffers used in transfection were sterilised by autoclaving or filtration with a 0.2 μ filter.

2 x HBS: (1 litre)

NaCl	280 mM
Hepes	50 mM
$Na_2HPO_4 12H_20$	1.5 mM

Adjust the solution to pH 7.08 (at room temperature) with about 6 ml of 1N NaOH.

DEAE-dextran: 10 mg/ml in PBS (Ca^{2+}, Mg^{2+}-free; GIBCO).

ß-Galactosidase assay 2 x buffer (Promega Corporation, Madison):

sodium phosphate, pH 7.3	200 mM
$MgCl_2$	2 mM
ß-mercaptoethanol	100 mM
ONPG*	1.33 mg/ml

*ONPG: o-nitrophenyl-ß-D-galactopyranoside.

2.7.2 Introduction of fused genes into mammalian cells by calcium phosphate co-precipitation

For preparation of the constructs containing the deleted 5' flanking sequence of the *CYP2A7A*, the DNA of pCAT2A7A5'-3.0 was double digested with *Hind*III and another enzyme which only digests the 3 kb insert of pCAT2A7A5'-3.0. Digested DNA fragments which contained a pCAT-basic vector and different sizes of 5' truncated insert were isolated, blunted and re-ligated to form the various constructs: pCAT2A7A5'-2.2 (*Hind*III-*Sst*I), pCAT2A7A5'-1.5 (*Hind*III-*EcoR*V), pCAT2A7A5'-

1.0 (*Hind*III-*Bgl*II), and pCAT2A7A5'-0.5 (*Hind*III-*Ava*I) (See chapter 3, Figure 3.12A).

The other series of constructs containing a 5' truncated promoter of *CYP2A7A* were obtained by digesting the plasmid pCAT2A7A5'-0.5 DNA with *Bal*31 as follows: 20 µg pCAT2A7A5'-0.5 DNA was cleaved with *Hind*III, and then treated with 5 units *Bal*31 (BRL) for 15-30 min. The aliquots were removed at different time points and instantly chilled on ice. All aliquots were combined, purified by phenol/chloroform extraction and ethanol precipitation. The pellet was redissolved in water, blunted using *Klenow* polymerase I. *BamH*I digestion was carried out to remove the part of the vector which was also deleted during the *Bal*31 digestion. The fragments containing the progressively deleted insert and the vector (the part from insert to *BamH*I site) were purified. These fragments were then ligated into the previously prepared part of pCAT-basic vector having a blunted *Hind*III end and a *BamH*I sticky end to form a series of constructs (See chapter 3, Figure 3.12B).

The cultured cells were seeded as described by Chen and Okayama (1987) at a density of 0.5-1.0 x 10^6 cells per 100-mm petri dish the day before transfection (See section 2.2.2). On the day of transfection, the cells were re-fed with fresh medium and continually cultured for 3 h. The DNA-calcium phosphate coprecipitate were prepared as follows: 10 µg of test plasmid DNA and 5 µg of the standard plasmid DNA (pSV-ß-Galactosidase plasmid, Promega Corporation, Madison) in a volume of 438 µl were mixed with 62 µl of 2 M $CaCl_2$ solution. The mixture was dropped slowly into another tube containing 500 µl of 2 x HBS solution with shaking and then allowed to stand at room temperature for 20 to 30 min. The calcium phosphate-DNA suspension was mixed gently and added directly to the medium on top of the cell monolayer. Four hours later, the cells were subjected to a 15% (v/v) glycerol/1 x HBS shock for 2 min. Cells were harvested 48 to 60 hr after transfection .

DNA-calcium phosphate coprecipitate: (for 100 mm dish)

DNA	X µl
H_20	to 438 µl
2M $CaCl_2$	62 µl
2 x HBS	500 µl

2.7.3 DEAE-dextran-mediated DNA transfection

DEAE-Dextran-mediated DNA transfection (Cullen, 1986) is a simple and efficient procedure which is specifically tailored for transient expression in COS cells. This protocol has been reported to yield transfection frequencies as high as 80%.

The COS cells were seeded at a density of 0.5 x 10^6 cells per 100-mm petri dish and incubated at 37^0C for overnight. Next day, the plate should be just subconfluent. Prepare transfection cocktail in a Eppendorf tube as follows: Add plasmid DNA into PBS solution (1.9 ml), vortex. Add 100 µl of DEAE-dextran into the DNA solution and mix. Aspirate culture medium from culture dish and rinse cells with 2 ml of PBS warmed to 37^0C and aspirate the PBS. Add transfection cocktail on top of the cell momolayer and distribute evenly by tilting plate. Incubate at 37^0C for 30 min with occasional gentle shaking to prevent drying. Add 5 ml of culture medium supplemented with 80 µM chloroquine and incubate at 37^0C for 2.5 h. Aspirate supernatant medium and replace with 3 ml culture medium containing 10% DMSO for 2.5 min. Remove the medium and add 10 ml of fresh medium. Incubate the cells at 37^0C for 48-60 h.

2.7.4 CAT activity assay

The CAT assay was carried out as follows ((Promega, protocols and applications guide): Transfected cells were washed 5 times with phosphate-buffered saline (PBS, Ca^{2+}, Mg^{2+}-free; GIBCO) and resuspended in 100 ul of 250 mM Tris-Cl, pH7.5. Cell lysates were prepared by freezing (-70^0C) and thawing (37^0C) for four cycles with vortexing after each freezing-thawing cycle, incubated at 60^0C for 10 min to inactivate

the endogenous acetylase and then centrifuged in a microfuge for 10 min. The reaction mixture is prepared as follows:

cell extract	50 µl
0.25M Tris-HCl, pH8.0	60 µl
(^{14}C) chloramphenicol	10 µl*
n-butyryl coenzyme A (5 mg/ml)	5 µl (final volume 125 µl)

* 1µl (^{14}C) chloramphenicol was diluted to 50 µl (0.5 µCi/ml)

Incubate the reaction at 37^{0}C for two hours. Terminate the reaction by adding 300 µl of xylene to each tube and shake by vortex for 30 seconds. Following centrifugation for 3 min, transfer 270 µl the upper phase (xylene) to a fresh tube. Add another 100 µl of fresh 0.25 M Tris buffer, pH 8.0, to wash the xylene phase by repeating the vortex and centrifugation. Carefully remove 200 µl of the upper, xylene phase and transfer it to a scintillation vial. Add 5 ml scintillation fluid and count the samples in a liquid scintillation counter.

For ß-galactosidase assay, add the following reaction mixtures directly to the wells of microtiter plates which has a maximum well volume of approximately 300 µl:

Positive control:

ß-galactosidase assay 2 x buffer	50 µl
ß-galactosidase	0.025 u*
water to final volume	100 µl

* diluted in water on ice from 1 u/1µl stock, prepared immediately before use.

Negative control:

ß-galactosidase assay 2x buffer	50 µl
non-transfected cell extract	30 µl
water to final volume	100 µl

Sample reaction:

ß-galactosidase assay 2x buffer	50 µl
transfected cell extract	30 µl
water to final volume	100 µl

Mix all samples by pipetting well contents up and down, and incubate the reaction at 37^{0}C for 2 h until a yellow colour is present. Stop the reaction by adding 150 µl of 1M sodium carbonate and mix by pipetting the contents of each well up and down. Absorption at 420 nm was measured.

2.8 Preparation and Analysis of RNA

2.8.1 Materials and buffers

4 M guanidinium solution (500 ml): 236.5 g guanidinium thiocyanate, 3.23 g (25 mM) NaCitrate, 0.5% (w/v) Sodium lauryl sarcosinate. Add deionized water to 500 ml and adjust pH to 7.0 with NaOH solution.

5.7 M CsCl solution (100 ml): 96 g CsCl, 37.2 g EDTA (0.1 M). Add ddH_2O to 100 ml and adjust pH to 7.0 with NaOH solution.

solution D: 4 M guanidinium thiocyanate, 25 mM sodium citrate, pH 7.0; 0.5% (v/v) sarcosyl and 0.1 mM 2-mercaptoethanol.

10 x MOPS (1000 ml): 40.86 g MOPES, 6.8 g Na-acetate 3 H_20 and 50 ml of 0.2 M EDTA. Add H_20 to 800 ml, adjust pH to 7.0 with NaOH solution and then add H_20 to 1000 ml. Keep the solution in cold (4 ^{0}C) and in dark place.

2.8.2 Guanidinium method for total RNA preparation from tissue culture cells

RNA preparation was carried as described by Ausubel *et al.* (1987). The cultured monolayer cells were washed twice, each time with 5 ml PBS. 5 ml 4 M guanidinium

solution was added to the cells (150 mm culture dish) and the cells should immediately lyse. The resultant extremely viscous solution was recovered by scraping the culture dish with a rubber policeman, transferred to a clean tube and homogenised with a whole glass homogeniser and a pestle (type A). The resulting homogenate was drawn up and down five times through a 20-G needle. It is critical that chromosomal DNA is sheared in this step in order to reduce viscosity and remove complete in the following centrifugation step. The cell lysate (5 ml) was carefully layered on top of CsCl cushion (5 ml 5.7 M CsCl) to create a step gradient and the interface should be visible. The sample was centrifuged at 35,000 rpm in a Beckman SW-55 rotor (or equivalent) 12 to 18 hr at 15^{0}C. after centrifugation the supernatant was carefully removed. The pellet was drained for 5 to 10 min, then resuspend it in 300 µl water and transferred to a clean tube. The RNA solution was extracted twice with same volume of phenol/chloroform, one time with chloroform/butanol and then precipitated with 1/10 volume of 3 M sodium acetate and 3 volume of cold ethanol at -20^{0}C for at least 1 hr. After centrifugation the RNA pellet was resuspended in a minimum volume of water and stored at -70^{0}C until required.

2.8.3 Single-step preparation of RNA from tissues and cells

The single-step method of RNA preparation (Chomczynski *et al* ,1987) allows isolation of RNA in 4 hr and provides both high yield and purity of undegraded RNA preparation. In addition, this procedure permits recovery of total RNA from small quantities of tissue as well as cultured cells

0.1 g frozen tissue was minced on ice and homogenised (at room temperature) with 1 ml of solution D in a glass-Teflon homogeniser and subsequently transferred to a clean tube. Sequentially, 0.1 ml of 2 M sodium acetate, pH 4, 1 ml of phenol (water saturated), and 0.2 ml of chloroform-isoamyl alcohol mixture (49:1) were added to the homogenate, with thorough mixing by inversion after the addition of each reagent. The final suspension was shaken vigorously for 10 seconds and cooled on ice for 15 min.

The mixture was centrifuged at 10,000g for 20 min at 4^0C. After centrifugation, RNA was present in the aqueous phase whereas DNA and proteins were present in the interface and phenol phase. The aqueous phase was transferred to a fresh tube, mixed with 1 ml of isopropanol, and then placed at -20^0C for at least 1 hr to precipitate RNA. Sedimentation at 10,000g for 20 min was performed and the resulting RNA pellet was dissolved in 0.3 ml of solution D, the solution was transferred into a 1.5 ml Eppendorf tube, and precipitated with 1 vol. of isopropanol at -20^0C for 1 hour. After centrifugation for 10 min at 4^0C the RNA pellet was washed with 75% ethanol, pelleted, dried, and dissolved in a minimum volume of water and stored at -70^0C until required. The purity of the RNA was further increased by repeating precipitation steps once or twice. Isopropanol precipitation can be replaced by precipitation with three volumes of ethanol.

2.8.4 Quantification of RNA

RNA concentration is determined like the DNA determination described in section 2.3.2. 3.3 µl of RNA solution was diluted to 1 ml with ddH_20 and mixed thoroughly. The absorption at 260 nm was read and the concentration was calculated according to the following equation:

$$RNA\ (\mu g) = A_{260} \times 10$$

The above equation is based on the assumption that 1 A_{260} = 33 µg RNA. A pure RNA sample has an A_{260}/A_{280} ratio of 2.0.

2.8.5 Primer extension analysis

Primer extension analysis (Ausubel, *et al.*, 1987) was carried out as follows: A 22-mer oligonucleotide was synthesised corresponding to positions 52-73 bp of the *CYP2A7* cDNA in the antisense orientation, 5'-GCCAGACAGACATCAAGACCAT-3'. Total RNA (50 µg) was coprecipitated with the (γ-^{32}P)ATP end-labelled oligonucleotide

prepared using T_4 polynucleotide kinase and the pellet was then resuspended in hybridisation buffer [80% (v/v) formamide, 40 mM PIPES, pH 6.4, 400 mM NaCl and 1 mM EDTA] and hybridised at 30^{0}C for overnight. The mixture was ethanol-precipitated and resuspended in 50 mM Tris-Cl, pH 8.3, 5 mM $MgCl_2$, 50 mM KCl, 10 mM dithiothreitol, 0.8 mM each of dNTPs, 0.5 unit RNasin and 40 units of avian myeloblastosis virus reverse transcriptase (BRL), in a reaction volume of 25 µl. Extension reaction was carried out for 1.5 hour at 42^{0}C, and terminated by adding 1 µl 0.5 M EDTA and 1 µl RNase A (1 mg/ml) at 37^{0}C for 30 min. Following ethanol-precipitation and resuspension in 8 µl of denaturing dye buffer, the reaction products were run on a 6% polyacrylamide, urea gel. Labelled products were sized by a standard DNA sequencing reaction.

2.8.6 Amplification of RNA (RT-PCR reaction)

(A) Reverse transcription reaction

The following reagents were assembled in a final volume of 20 µl: 2 µl of 10 x PCR buffer (Promega Ltd.), 10 µl of 2 mM each dNTP solution, 0.5 µl of RNasin (40 unit/µl, Promega Ltd.), 1 µl of oligo$(dT)_{6-8}$ solution (0.1 µg/ µl), 1 µl of 100 mM $MgCl_2$ solution, 1-5 µl of total RNA sample (5-10 µg), and 200 to 400 units of MoMuLV reverse transcriptase. The reaction was incubated 15 min at 23^{0}C, then 60 min at 43^{0}C, and transferred to 95^{0}C water bath for 10 min. After heat treatment the reaction was quick-chill on ice.

(B) PCR amplification

To the heat-treated reverse transcription reaction, 80 µl of 1 x PCR buffer was added containing 10 to 50 pmol each of upstream and downstream primer and 2.5 units of *Taq* polymerase. Then 100 µl of mineral oil was added on the top of the PCR solution to prevent evaporation of liquid during thermal cycling. A thermal cycle profile was: (1) denaturing for 1 minute at 94^{0}C, (2) annealing primers for 1 minute at 55^{0}C, (3)

extending the primers for 2 min at 72^{0}C. After 30 to 35 cycles the reaction was incubated at 72^{0}C for 8 min, then 5 to 10 µl was used for analysis in a 1% (w/v) agarose gel.

2.8.7 Northern blot and hybridisation analysis

The following mixture should be prepared for each RNA sample to be analysed (Ausubel, *et al.*, 1987):

formamide	4 µl
formaldehyde	1.6 µl
10 x MOPS	1.2 µl
RNA (10-15 µg)	3.2 µl

Mix the samples well by vortexing, and incubate at 65^{0}C for 20 min. Add 2 µl loading buffer (same as that used in DNA electrophoresis) to each sample and load the samples along with 3 µl of an RNA ladder (0.24-9.5 kb derived from bacteriophage T7, yeast 2 µ circle and bacteriophage lambda DNA. BRL, Paisley, UK.) onto gel. Run the gel in 1x MOPS at a constant voltage of 5 V/cm for 3-5 h. The transfer of the RNA from gel to a Hybond-N filter was same to the procedure in Southern blot (Section 2.3.6) except that there was no denaturing and neutralising treatment.

After the transfer was completed, the filter was placed in hybridisation mixture [50% (v/v) formamide, 5 x SSC, 5 x Denhardt's solution, 0.1% (w/v) SDS, 50 mM sodium phosphate, pH 6.5 and 50-100 µg/ml denatured salmon sperm DNA] for at least 4 h at 42^{0}C in a glass tube in a preheated oven. Probe, prepared and radiolabelled as described in Sections 2.3.4 and 2.3.7, was added directly to the hybridisation solution following denaturation. The reaction was incubated for 16-18 h at 42^{0}C. The filter was washed 3 times in 2 x SSC, 0.1% (w/v) SDS. Higher stringency washes of 0.2 x SSC, 0.1% (w/v) SDS at room temperature were carried out for 15 min. After washing the filter was wrapped in Saran wrap and exposed to Kodak X-Omat AR-5 film at -70^{0}C in a cassette with intensifying screens.

2.9 Analysis of Proteins

2.9.1 Protein estimation

Protein concentration was determined spectrophotometrically according to Lowry *et al* (1951), using serial dilution of bovine serum albumin (BSA) to generate a standard curve for each assay. The absorption obtained at 600 nm was linear over the range 0-200 µg/ml, and all samples were therefore diluted (0.1 M NaOH) to lie within this range.

2.9.2 Denaturing gel electrophoresis (SDS-PAGE)

SDS-PAGE gel electrophoresis was carried out according to the method of Laemmli (1970). All protein samples were analysed by SDS-PAGE prior to immunoblotting to demonstrate equivalence of loading and to ensure that the samples were not degraded. Samples were prepared for electrophoresis by dilution in distilled deionised water to a concentration of 3 mg/ml of total protein, before the addition of an equal volume of "boiling mix" (0.05 M Tris-HCl pH 8.0, 2% (w/v) SDS, 5% (v/v) ß-mercaptoethanol, 10% (v/v) glycerol, 0.005% (w/v) bromphenol blue) to give a final protein concentration of 1.5 mg/ml. Proteins were denatured by heating to 100 ^{0}C for 5 min. For Coomassie Brilliant Blue stained gels, 15 µg of total cellular protein was loaded per track. Molecular weight markers containing alpha-lactalbumin (Mr=14200), soybean trypsin inhibitor (Mr=20100), trypsinogen (Mr=24000), bovine erythrocyte carbonic anhydrase (Mr=29000), rabbit muscle glyceraldehyde-3-phosphate dehydrogenase (Mr=36000), chicken egg albumin (Mr=45000) and bovine serum albumin (Mr=66000) were prepared according to the manufacturer's instructions and were loaded at 10 µl per track. SDS-PAGE gel (0.15 cm) was cast using Biorad Protean II electrophoresis apparatus. Separating gel composition was dependent on the particular protein, for example P4502A proteins were separated by electrophoresis in a 10% gel. The gel composition is given as below:

10% Separating gel: 9.65 ml 40% (w/v) acrylamide, 6.65 ml 2% (w/v) bis-acrylamide, 10 ml 1.5 M Tris (pH8.8), 0.8 ml 10% (w/v) SDS, 13 ml ddH_2O, 400 µl 10% (w/v) APS, 20 µl TEMED.

4.5% Stacking gel: 1.23 ml 40% (w/v) acrylamide, 0.85 ml 2% (w/v) bisacrylamide, 1.25 ml 1.0 M Tris-Cl (pH6.8), 0.1 ml 10% (w/v) SDS, 6.4 ml dH_2O, 100µl 10% (w/v) APS, 10 µl TEMED.

Gels were poured to a height of 12 cm and immediately overlaid with the water saturated isobutanol. The rapid separation of this mixture into organic and aqueous phases resulted in an even boundary forming at the interface with the gel. Once set, the butanol overlay was removed and the gel rinsed several times with dH_2O, before a stacking gel was poured to the top of the separating gel and a well-forming comb (15 tracks) inserted. When the stacking gel had set, the plate assemblies were securely attached to the central cooling reservoir of the Protean II apparatus and the position of the wells clearly marked. The top buffer chamber was filled with freshly prepared electrophoresis buffer [5 mM Tris, 50 mM glycine, pH 8.3 and 0.02% (w/v) SDS], and the combs carefully removed. Samples were loaded before the entire apparatus was placed in a gel tank containing 3 litres of electrophoresis buffer. Gels were run (60 mA/gel stacking, 30 mA/gel separating) with cold water cooling and with continuous mixing of the electrophoresis buffer to minimise any build-up of buffer gradients. When the dye front was approximately 1 cm from the bottom of the plates, electrophoresis was stopped, and the gels removed from the glass plates. The gel can be stained with Coomassie Brilliant Blue , or used for a western blot.

(A) Staining SDS-PAGE gel with coomassie brilliant blue

The gel was immersed in staining solution [0.25% (w/v) Coomassie Brilliant Blue in 45.5% (v/v) methanol, 9% (v/v) acetic acid] with slow shaking for 4 h at room

temperature, then destained in the solution of 10% (v/v) methanol, 9% (v/v) acetic acid to visualise the protein bands.

(B) Immunoblotting (Western blotting)

Immunoblotting was performed essentially as described by Lewis *et al.* (1988). After electrophoresis, the gel was removed from the glass plates, and then a gel/membrane "sandwich" was assembled as follows: A large basin was filled with transfer buffer [20 mM sodium orthophosphate 12 H_2O, 20% (v/v) methanol], into which was placed a plastic cassette containing a nylon "Scotchbrite" pad, a sheet of 3 mm filter paper cut to a size slightly larger than the gel was put on top of the nitrocellulose filter, covered with another sheet of filter paper, a further nylon pad and the cassette were closed, thus clamping the gel sandwich firmly together. Each gel was assembled in a similar manner, and the cassette(s) was placed in a Biorad transblot apparatus containing freshly prepared transfer buffer. Proteins were then transferred electrophoretically (250 mA, overnight) from the gel to the nitrocellulose sheet. Following transfer, the nitrocellulose filter was cut to an exact replica of the gel and was placed in a plastic container containing sufficient TBST [0.9% (w/v) NaCl, 0.6% (w/v) Tris-Cl, pH 7.9 and 0.05% (v/v) Tween-20] to cover the filter. The filter was washed twice with TBST for 5 min, and then incubated with the blocking solution [5% (w/v) low-fat dried milk in TBST] for 1-2 h. The filter was further washed with TBST (1 x 15 min, 2 x 5 min), before being incubated with the first antibody (rabbit anti-CYP2A antibody, diluted 1:500 to 1:1000 in TBST) for 1 hour. After washing with TBST (3 x 10 min) to remove excess unbound first antibody, the filter was placed in the HRP-labelled second antibody (donkey anti-rabbit IgG-HRP for monoclonal primary, diluted 1:1000 in TBST) for 1 hour. After washing with TBST (3 x 10 min), freshly prepared substrate solution [0.05% (w/v) 4-chloro-1-naphthol, 17% (v/v) methanol, 0.01% (v/v) hydrogen peroxide in TBS] was added, and the immunoreactive polypeptides were visualised by the appearance of purple bands on the filter. For higher sensitivity, the

filter was washed in distilled water for 15 min with shaking and then placed in 50 ml TBST containing 50 μCi (0.19 MBq) ^{125}I-protein A (Amersham International, plc). After 30-60 min, the filter was exhaustively washed with TBST and subsequently exposed (Kodak X-Omat AR5 X-ray film) 1-3 days with intensifying screens at -70^{0}C in an autoradiography cassette. When the amount of the protein of interest was particularly low, enhanced chemiluminescent detection (ECL) was used (Amersham International, plc). This technique relies on the generation of an unstable photodynamic complex on reaction of H_2O_2, HRP-labelled secondary antibody and the cyclic diacylhydrazine luminol. Dissociation of this complex results in luminol returning to its ground state with associated light emission. The amount of light emitted is directly proportional to the amount of immobilised protein bound to the HRP-labelled secondary antibody.

CHAPTER 3: ISOLATION AND CHARACTERISATION OF THE HUMAN *CYP2A7A* GENE

3.1 The Background of *CYP2* Gene Family: Structure and Regulation

3.1.1. The methods used for isolation of CYP genes

Since DNA was found as the active genetic principle, studies in the field of molecular biology have made rapid progress catalysed by many theoretical and technical advances. During this period the technique of DNA cloning also developed rapidly, and it became possible to construct genomic DNA or cDNA libraries in the late 1970s when Maniatis *et al.* (1978) devised a strategy to generate and clone large numbers of random fragments of mammalian genomic DNA. This progress gave a vigorous push to the understanding of cytochrome P450s at the molecular level. More than 220 genes of cytochrome P450 have been isolated during the last decade. The strategy of isolating and characterising P450 genes is illustrated in Figure 3.1.

Briefly, RNA is isolated from the appropriate tissue, usually from liver in which there are high levels of P450 except for those with a tissue-specific manner of expression. The double-stranded cDNA is synthesised from the isolated mRNA and inserted into a vector which allows the inserted cDNA to generate a protein product. The protein can be detected with a specific antibody, or the clone containing an interesting cDNA fragment can be isolated from a population of thousands of diverse cDNAs by screening with an oligonucleotide probe.

After the cDNA has been isolated, it can be used as a probe to screen a genomic DNA library to get its corresponding genomic clone. The genomic library is constructed from total DNA isolated from cultured cells or tissues. Total DNA is partially digested with a restriction enzyme and then fragments ranging from 15-20 kb in size are inserted into a

Figure 3.1 The basic strategy of isolating a P450 gene.

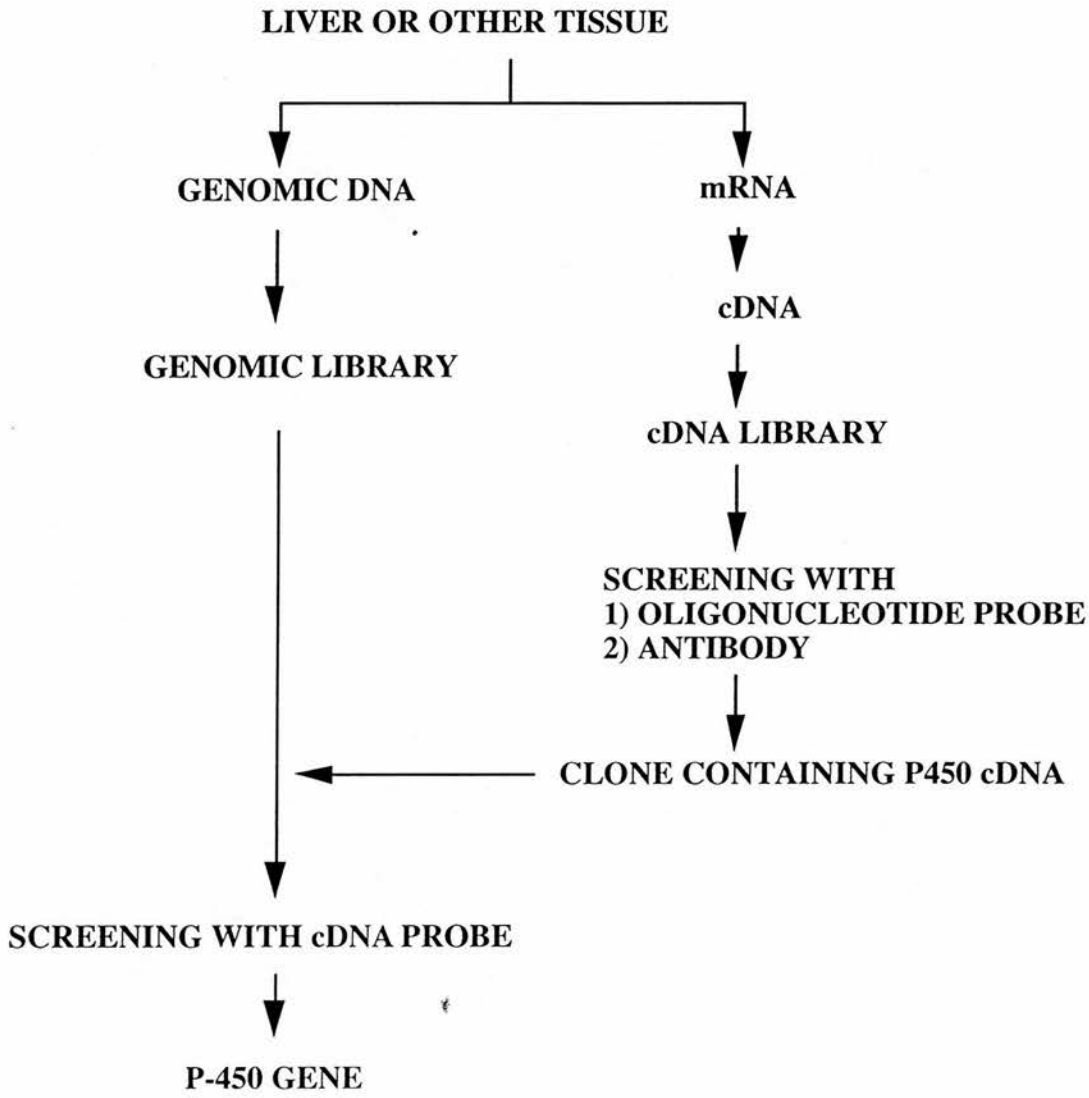
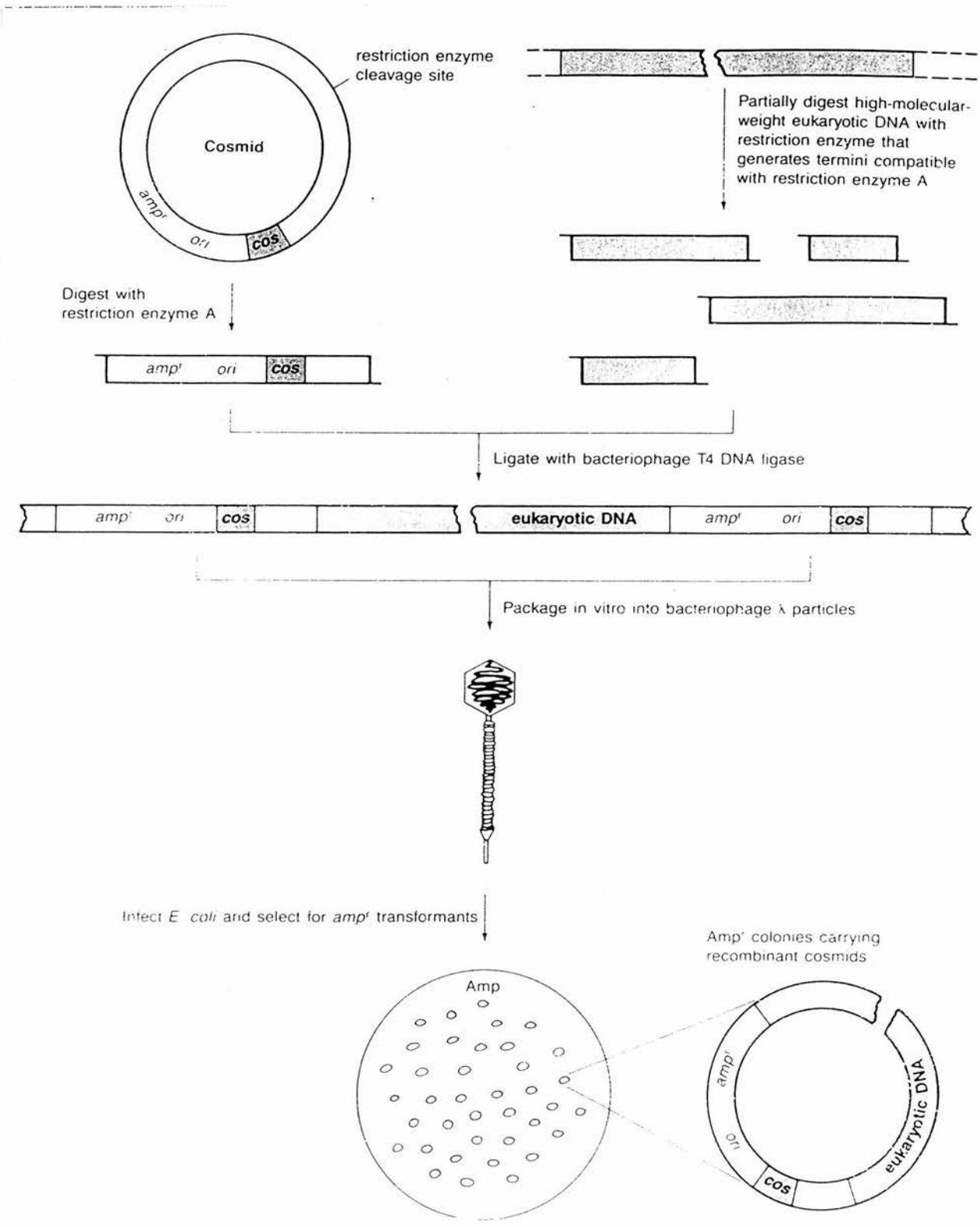


Figure 3.2 The steps involved in cloning in cosmid vectors (Maniatis *et al.*, 1989)



Lambda phage cloning vector, or fragments ranging from 30-45 kb in size are inserted into a cosmid vector which is suitable to isolate a full length of gene in one clone (Figure 3.2).

The cDNA and genomic DNA clones of P450 can be used in a variety of studies in pharmacology and molecular biology. The cDNA can be used to determine the amino acid sequence of an enzyme. The human P450 can be heterologously expressed in E. coli or yeast cells to obtain a protein for further studies. The genomic DNA can be used to determine the structure of a P450 gene and to identify the important regulatory domains which can be used to isolate and characterise receptors and factors required for control of gene expression.

3.1.2 The structural features of CYP2A subfamily genes

The *CYP2A* subfamily has been extensively studied in the rat and the mouse. Three genes of rat *CYP2A* subfamily have been isolated and sequenced recently, and the genomic structures of the *CYP2A1* and *CYP2A2* genes have been described (Matsunaga *et al.*, 1990). These genes contain nine exons, in common with other *CYP2* family genes. The size of the *CYP2A1* gene is almost twice as long as *CYP2A2* or *CYP2A3* due to a 14 kb fifth intron in *CYP2A1*. The 5'-flanking region of *CYP2A1* contains a typical TATA box at -27 to -24 bp and a reverse CCAAT box at -85 to -90 bp upstream of the transcription start site. The *CYP2A2* gene also has a typical TATA box at -24 bp, but contains no CCAAT box. *In vitro* transcription of the *CYP2A1* and *CYP2A2*, both genes were accurately transcribed in extracts prepared from livers of male and female rats. This result is surprising in view of the fact that the *CYP2A1* is expressed only in adult female rats while the *CYP2A2* is expressed in adult males (Nagada *et al*, 1987).

Two genes of the mouse *Cyp2a* subfamily, designated *Cyp2a-4* and *Cyp2a-5*, have been isolated and sequenced (Lindberg *et al.*, 1989). The comparison between the two

Table 3.1 The genes in *CYP2A* subfamily and their chromosomal location

Species	cDNA symbol	Chromosomal location	Reference
Rat	*CYP2A1*		Nagata *et al.*, 1987
	CYP2A2		Matsunaga *et al.*, 1990
	CYP2A3		Ueno and Gonzalez 1990
Mouse	*Cyp2a-4*	7	Miles *et al.*, 1990
			Lindberg *et al.*, 1989a
	Cyp2a-5	7	Lindberg *et al.*, 1989b
			Lange *et al.*, 1990
Human	*CYP2A6*	19q13.1-13.2	Miles *et al.*, 1989a; 1990
			Yamano *et al.*, 1990
	CYP2A7		Yamano *et al.*, 1990
Hamster	*CYP2A8*		Fukuhama *et al.*, 1989
			Koga *et al.*, 1990
	CYP2A9		Lai and Chiang 1990
Rabbit	*CYP2A10*		Nelson *et al.*, 1993
	CYP2A11		Nelson *et al.*, 1993

genes, each about 8 kb in size, exhibits virtually identical exon-intron organisation and over 96% nucleotide sequence homology, including both 5'- and 3'-regions. Both genes have TATA boxes at 30 bp and CAAT boxes at 100 bp upstream from the transcription start site. The *Cyp2a-4* is expressed predominately in female liver, male kidney, and the *Cyp2a-5* is mainly activated in female liver and kidney (Squires *et al.*, 1988). In order to gain some insight into the sex-specific transcriptional of these Cyp2a genes, the mRNAs from these different tissues were used as templates to determine transcription start site. The results demonstrated that there was not sex-specific transcriptional start site and both genes have similar multiple start sites in all tested tissues even after digestion with the highest tested amount of mung bean nuclease (400 units). The high amount of nuclease was used to avoid incomplete digestion. Two major start sites were at 24 bp and 14 bp upstream from the initiation codon, respectively.

Up to date, there is no report about the structure of the human *CYP2A* gene. It was the aim of this study to characterise structure and regulation of the human *CYP2A* genes.

3.1.3 The multiple control of CYP2 gene expression

Eukaryotic genes are under the control of transcription factors which bind to promoter/enhancer elements to regulate transcription. Some of these factors are active only in the presence of an inducing stimulus or in a specific tissue, thereby producing a specific pattern of inducible or tissue-specific gene expression. In addition, some factors are constitutively active binding to specific upstream sequences. The binding of these factors and their interaction with basal transcriptional complex result in normal levels of transcription in all tissues. In the absence of such factors, therefore, the basal transcriptional complex can produce only a very low level of transcription.

The expression of genes in *CYP2* family is under the control of different regulation mechanisms, probably including receptors which are responsible for the induction of

foreign chemical and some tissue-enriched transcriptional factors that control constitutive expression. Developmental and sex-dependent regulations have also been described (Gonzalez, 1990). Although a suitable cell culture system is lacking for defining the regulatory elements of the inducible *CYP2* genes, especially the human inducible *CYP2* genes, recent studies on rodent *CYP2* genes and bacterial *CYP102* genes have made some progress for understanding *cis*-regulatory mechanism(s). Several models for the expression of these genes are reviewed below:

(A) Induction response of *CYP2* genes

Of great interest are the mechanisms by which large numbers of foreign chemicals can induce specific *CYP* genes to increase transcription. A well studied case is induction by barbiturates. A number of genes in *CYP2* and *CYP3A* subfamilies are induced by phenobarbital and the induction occurs at the level of transcriptional activation, involving an increase in the synthesis of mRNA and in the rate of transcription initiation (Omiecinski *et al.*, 1985; Honkakoski *et al.*, 1992; Hahn *et al.*, 1991). The two major phenobarbital-inducible forms are rat *CYP2B1* and *CYP2B2* genes (Hardwick *et al.*, 1983). Phenobarbital administration to rat causes a dramatic increase in the level of CYP2B mRNA (Omiecinski *et al.*, 1986). Another phenobarbital-inducible gene is rabbit *CYP2C* which is induced by a similar mechanism as *CYP2B* (Zhao *et al.*, 1990).

Unfortunately, to date, the kind of factors or receptors are involved in the phenobarbital induced expression of *CYP2* genes in mammals still remains to be elucidated because expression is not phenobarbital-inducible in cultured cell lines. However, progress has been made recently in the bacterium *Bacillus megaterium* (He and Fulco, 1991). *CYP102* ($P450_{BM3}$) is one of the barbiturate inducible genes in this bacterium which incorporates both P450 and P450 reductase in a single soluble 119-KDa polypeptide. CYP102 is involved in the metabolism of fatty acids, and the mechanism of barbiturate inducible expression of *CYP102* has been partially characterised (Shaw and Fulco, 1993). Comparing the sequences between the *CYP102* gene and other barbiturate

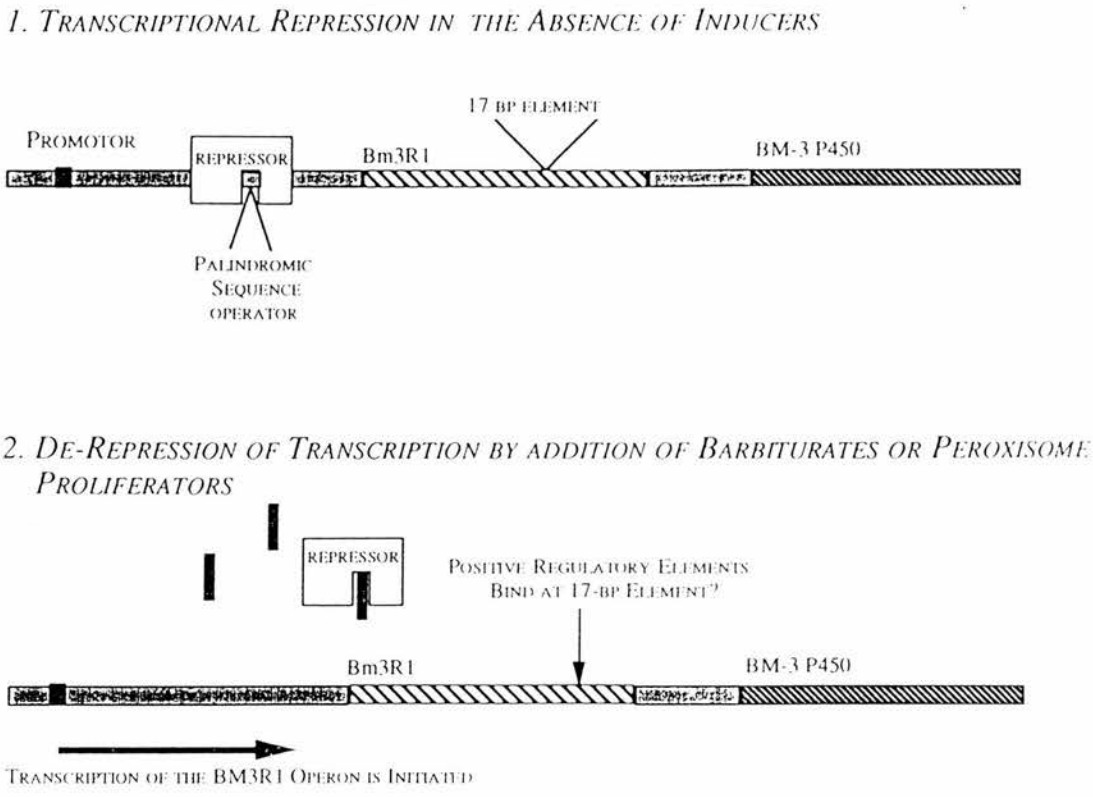
inducible genes, a 15 bp element, ATCAAAAGCTGGAGG, was found existing in the rat *CYP2B1* and *CYP2B2* genes. A protein in the extracts prepared from normal *Bacillus megaterium* cells binds strongly to the 15 bp element, but this binding is dramatically blocked with protein extracts from phenobarbital treated cells. Conversely, the 15 bp homologues in rat *CYP2B1* and *CYP2B2* genes are weakly bound by a protein in untreated nuclear extracts of rat, but the binding is much stronger with the protein from phenobarbital treated rats, suggesting a different regulation mechanism with that of bacterial *CYP102* gene.

The protein binding to the 15 bp element in *Bacillus megaterium* is identified as the Bm3R1 repressor, and the gene encoding the protein is located upstream of the *CYP102* gene (Shaw and Fulco, 1993). The Bm3R1 repressor binds specifically to a palindromic 20-bp site in the promoter-operator region of the Bm3R1 gene to inhibit transcription of the gene. The binding between Bm3R1 repressor and its operator, identified by *in vitro* gel retardation studies, is strongly blocked by addition of 2 mM phenobarbital, but not by addition of non-inducer. However, it is still unclear whether the phenobarbital acts directly as depicted in Figure 3.3 or indirectly to release the repressor from the operator.

Figure 3.3 shows the elements which have been proposed for this regulation. A palindromic 20-bp site is located upstream of the open reading frame of Bm3R1 and interacts with the Bm3R1 protein. It is believed that Bm3R1 repressor binds tightly to this motif to inhibit transcriptional initiation.

In addition, rat *CYP2B1* and *CYP2B2* are also regulated by a synthetic glucocorticoid dexamethasone (DEX). After DEX injection, rat CYP2B1/CYP2B2 mRNAs increased twelve fold (Simmons *et al.*, 1987). Jaiswal *et al.*, (1990) inserted the 5'-flanking segment of *CYP2B2* gene into the vector pBLCAT which contains a 'basal' thymidine kinase gene promoter (including a TATA element, but no other upstream promoter elements) and a reporter gene CAT. The expression of CAT is under the control of this

Figure 3.3 Model of transcriptional activation of the *CYP102* gene. Schematic diagram of barbiturate and peroxisome proliferator-mediated induction in *Bacillus megaterium*. In the absence of inducer, the repressor (Bm3R1) binds to the operator, and prevents initiation of transcription. Barbiturate or peroxisome proliferator interact with Bm3R1 and inhibit its binding to the opreator. Transcription is then initiated with formation of a bi-cistronic message encoding Bm3R1 and CYP102 (Adapted from English *et al.*, 1994).



fused promoter-enhancer segment and the expression level of CAT can be determined in transfected cells. Using this technique, it was found that transfected rat hepatoma H4II cells show only a low level CAT expression in absence of DEX. Addition of DEX leads to an approximately 20 fold increase in CAT expression. The sequence analysis of 5'-flanking regions shows that there is a putative glucocorticoid response element (GRE) in the rat *CYP2B2* gene and in the rat *CYP2B1* gene as well (Ding and Wolf, unpublished result). The GRE is clustered together with CACCC and CCAAT binding sites (Figure 3.4A). It has been found that duplication of GRE, or the combination of GRE with either a CACCC, CCAAT or Sp1 motif increases the transcriptional activity in responding to steroid hormone induction, while GRE itself mediates a very low induction upon treatment with dexamethasone (Strähle *et al.*, 1989; Jantzen *et al.*, 1987; Schüle and Muller, 1988). A study on the cooperation of GRE with other transcription factor binding sites is shown in Figure 3.4B. All of these results suggest that the GRE in rat *CYP2B1/2B2* can function as an enhancer and may cooperate with the CACCC or CCAAT motif to increase heterologous thymidine kinase gene promoter activity in cells transfected by a *CYP2B1*/TK fusion gene.

(B) Constitutive expression of *CYP2* genes

Many *CYP* genes are constitutively expressed in the absence of exogenous inducers. Although extrahepatic tissues, such as lung, kidney, intestine and skin, have lower levels of expression, the basal activity of *CYP* gene is mainly in liver. The liver-specific gene expression is due to factors which are enriched in liver tissue. Hepatocyte transcriptional factors have been extensively studied (Lai and Darnell, 1991; Kuo *et al.*, 1992), and eleven genes (Sladek *et al.*, 1990; Gonzalez *et al.*, 1993) encoding positively-acting transcription factors have been cloned. These factors, which have limited tissue distribution and appear to regulate many important liver-specific genes, are all members of gene families: C/EBP, HNF-1, HNF-3, HNF-4 and albumin gene D region-binding protein (DBP) (Johnson *et al.*, 1987; Costa *et al.*, 1989; Frain *et al.*,

Figure 3.4

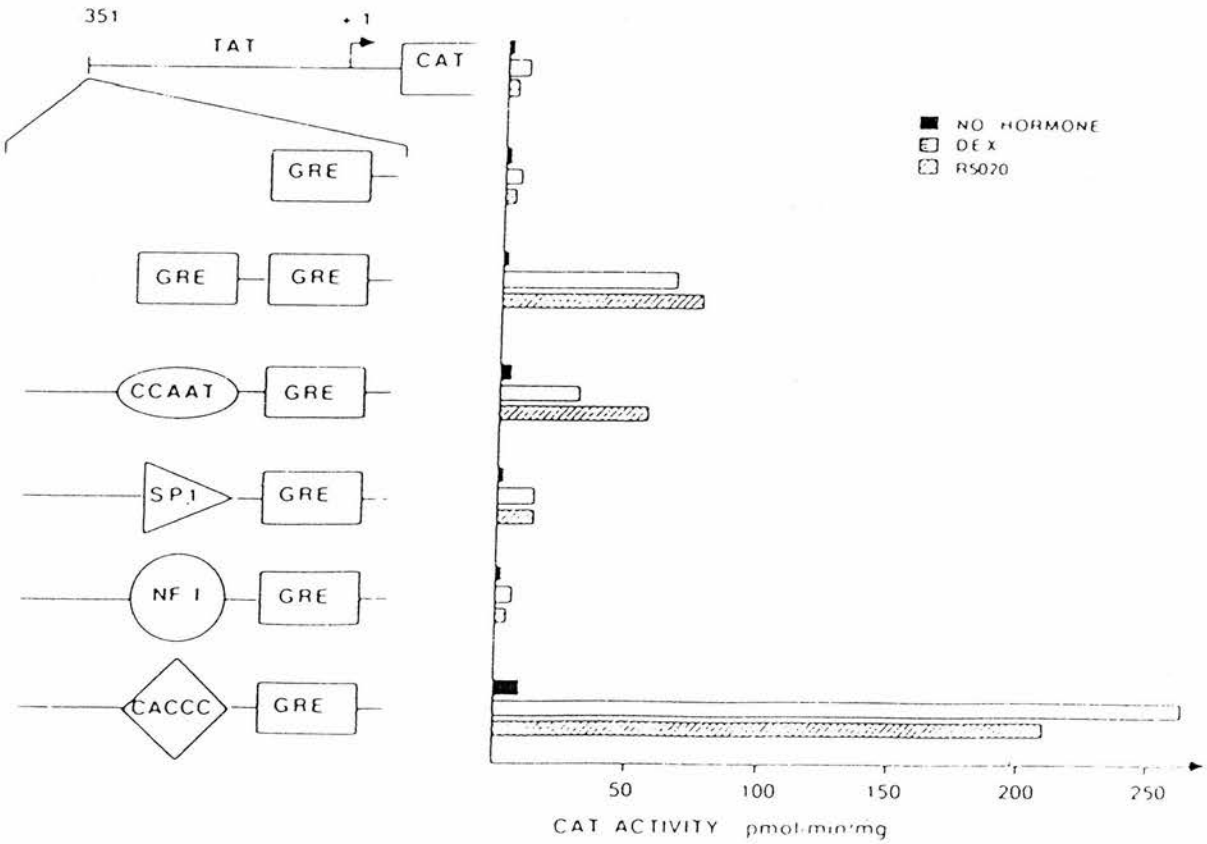
A: The sequence of rat *CYP2B1/2B2*-GRE (Jaiswal *et al.*, 1990).

5'-$^{-1376}$CCACC<u>CCAAT</u>AATATCAGTTAGGGTACAAAGTGTTCAAAC^{-1346}-3'

GRE Consensus GGTACANNNTGTTCT

The CACCC-box is in bold letters and the CCAAT is underlined.

B: Cooperation of PRE/GRE with CACCC-box and other promoter elements (Strähle *et al.*, 1989; Jantzen *et al.*, 1987; Schüle *et al.*, 1988b). The plasmids containing several binding site combinations were transfected into recipient cells. The activities induced by glucocorticoid were analysed for CAT activity.



1989; Baumhueter *et al.*, 1990). Tables 3.2 (Lai and Darnell, 1991) and 3.3 (Sladek and Darnell, 1992) summarise tissue distribution and some basic information on the members of these transcription factor families.

In an *in vitro* transcription assay with cell-free liver extract as well as in a DNA binding study, Ueno and Gonzalez (1990) found a positive control element located between -127 and -89 bp of the rat *CYP2E1* promoter. Deletion of the element results in a 90% decrease of the *in vitro* transcriptional activity of adult liver extract. This *cis*-acting element exhibits significant sequence similarity with the core HNF-1 binding motif which is involved in regulation of other liver-specific genes. The DNA binding results also reveal that this element can be bound by HNF-1, suggesting that rat *CYP2E1* gene is positively regulated in adult rats by HNF-1 or a HNF-1 related factor. However, CYP2E1 gene is not expressed in cultured cells, despite the presence of HNF-1 mRNA and protein (Baumhueter *et al.*, 1988). This demonstrates that the liver-enriched transcription factor HNF-1 is not sufficient for expression of *CYP2E1*, other factors which do not directly bind to the DNA may also be involved in the regulation of *CYP2E1* expression.

The transcription of *CYP2C6* gene is maximally activated when male and female rats reach puberty, and this activation coincides with the expression of the DBP (albumin gene D region binding protein) which is believed to control albumin gene expression in adult rats at a high level. Yano *et al.* (1992) found that DBP can activate transcription of *CYP2C6* by binding to a promoter element of the gene, and this binding affinity is 17-fold higher than to the albumin promoter. Both of the DBP-binding regions of albumin and *CYP2C6* genes have a size of 20 base pairs, but have no sequence similarity. Another factor (C/EBPa, a CAATT/enhancer binding protein) can bind to the DBP-binding regions of the albumin and *CYP2C6* , but only activates transcription of the former gene (Yano *et al.*, 1992).

Table 3.2 Hepatocyte transcription factors required for optimal expression of several tissue-specific genes.

Transcription factor family	Target genes in liver	DNA-binding motif
C/EBP C/EBPα C/EBPβ (NF-IL6, LAP, IL-6DBP)	Alb, TTR, α1 AT	bZIP
HNF-1 HNF-1α (LF-B1, APF) HNF-1β (vHNF-1)	α and βFib, α1-AT, Alb, TTR, AFP	POU-homeodomain
HNF-3 HNF-3α HNF-3β HNF-3γ	TTR, α1-AT, Alb AFP, TAT, PEPCK	ND
HNF-4	TTR, apo CIII, α1-AT PyK, GS	Zinc finger, nuclear receptor (steroid–thyroid family)

Three of the four families belong to classes of factors that share a known DNA-binding structure. C/EBP (CCAAT/enhancer-binding protein), HNF-1,3,4, hepatocyte nuclear factor-1,3,4; bZIP, basic leucine zipper; AFP, α-fetoprotein; α1-AT, α-1 antitrypsin; apo CIII, apolipoprotein CIII; Alb, albumin; Fib, fibrinogen; GS, glutamine synthetase; PEPCK, phosphoenol pyruvate carboxykinase; PyK, pyruvate kinase; TAT, tyrosine amino-transferase; ND, not determined. The listing of representative target genes is not meant to be comprehensive since each factor is known to have binding sites in at least ten gene loci.

Table 3.3 Tissure distribution of hepatocyte transcription factors

Transcription factor	Tissue distribution											
	Endodermal origin											
	Liver	Intestine	Lung	Stomach	Kidney	Brain	Spleen	Heart	Thymus	Fat	Ovary	Testes
HNF-1α	+ +	+ +	−	+	+ +	−	±	−	−	ND	−	ND
HNF-1β	+(−)	+	+	+	+ +	−	−	−	−	ND	+	ND
C/EBPα	+ +	+	+	ND	+(−)	+	±	±	±	+ +	ND	−
C/EBPβ	+ +	+ +	+ +	±	+(−)	±	+(−)	+(−)	±	+ +	+	+(−)
C/EBPδ	+	+ +	+ +	ND	+	±	+	±	ND	+	ND	ND
C/EBPε	−	−	−	−	−	−	−	−	−	−	−	ND
HNF-3α	+ +	+	+	+	−	−	−	ND	ND	ND	−	−
HNF-3β	+ +	+	+	+	−	−	−	ND	ND	ND	+	−
HNF-3γ	+ +	+	−	+	−	−	−	ND	ND	ND	+	+
HNF-4	+ +	+ +	−	−	+ +	−	−	−	ND	−	ND	ND

Shown are the relative amounts of the mRNA levels based on northern blot or T2 ribonuclease analysis: + +, most abundant; +, moderate; ±, marginal; −, absent; ND, not determined. Examples of protein levels that differ from the RNA levels are given in parentheses.

In rabbits, five genes of the *CYP2C* subfamily have been sequenced (Zhao *et al.*, 1990; Chan and Kemper, 1990; Pendurthi *et al.*, 1990). Study (Venepally *et al.*, 1992) of rabbit *CYP2C1/CYP2C2* genes shows that several promoter elements including binding sites of liver-enriched factors such as HNF-1 and HNF-4 are present on the 5'-flanking region of *CYP2C1/CYP2C2* genes. In order to characterise the functional significance of these regulatory elements, the 5'-flanking regions of both genes were progressively truncated from the 5' end and subcloned upstream of the CAT reporter gene. The promoter activities were analysed by transfecting these fused constructs into human hepatoma cells (HepG2) and monkey kidney cells (COS-1), separately. The results showed that the promoter activity in COS-1 cells was much lower than that in HepG2 cells. Deletion of the sequence from -116 to -67 resulted in a 90% reduction of promoter activity in HepG2, but had little effect on the promoter activity in COS-1 cells. In a further gel retardation study, one protein has been identified in the nuclear extracts of HepG2 cells that can bind to the sequence between -116 and -67, but there was not a similar protein in COS-1 cells. These results demonstrated that a hepatic-specific factor functions in the regulation of *CYP2C1/CYP2C2* genes. This conserved protein binding sequence, designated as HPF1 (HepG2-specific P450 2C factor-1, Venepally *et al.*, 1992), has been found in the genes of rodent *CYP2A*, *CYP2C* and *CYP2D* subfamilies and also been found in the human *CYP2A7A* gene in my study (For details, see Section 3.3.4).

(C) Sex-specific expression of rodent *CYP2* genes

The sex-specific expression of the *CYP2A* genes varies among different rodent models and has not been found in humans. The best studied sex-specific *CYP2* genes are male rat specific *CYP2C11* and female-specific *CYP2C12* (Ryan *et al*, 1982; 1984). The enzymatic activity of the CYP2C11 is only present in adult male rats, and is dependent on androgen exposure both during the neonatal and adulthood stage of development (Gonzalez, 1989). The Northern hybridisation with a *CYP2C11* specific

oligonucleotide probe establishes that CYP2C11 mRNA is exclusively expressed in the liver of adult male rats and is absent in female liver (Morishima *et al.*, 1987; Zaphiropoulos *et al.*, 1988). The high level of expression probably results from transcriptional activation of this gene.

Rat *CYP2A1/CYP2A2* genes are also sex-specifically regulated (Matsunaga *et al.*, 1988). The *CYP2A1* gene is activated soon after birth in both males and females and is specifically suppressed in postpubertal males but remains active in females (Nagata *et al.*, 1987). In contrast, the *CYP2A2* gene is activated when males reach puberty but is inactive through the life of females (Matsunaga *et al.*, 1988). To determine the mechanisms of sex-specific regulation, both rat genes have been isolated and sequenced (Matsunaga *et al.*, 1990). Both genes have nine exons that display 93% nucleotide similarity. *In vitro* transcription studies were carried out using two plasmid constructs as templates, each of them contained -6 kb to +1.5 kb from the transcription start site of the rat *CYP2A1* or *CYP2A2* genes, separately. The constructs were transcribed in an *in vitro* system with nuclear extracts derived from adult male liver, or female liver. The results showed that there was not a sex-specific transcription site. This study, therefore, did not explain the *in vivo* sex-specific activities of *CYP2A* rat genes (Matsunaga *et al.*, 1990).

Lindberg *et al.* (1989) recently described the structures of the mouse *Cyp2a* genes. In spite of their high sequence homology, the transcriptions of *Cyp2a4* and *Cyp2a5* are regulated differently. Both genes are expressed in female liver, whereas only *Cyp2a4* is expressed in castrated male liver. Mung bean nuclease digestion assays reveal that there is not a sex-specific or tissue-specific transcription start site (See Section 3.3.3).

Taken together, the investigation on sex-specific expression of the *CYP2A* gene, including the gene structure and the transcription start site failed to show a significant difference between the male-specific and female-specific *CYP2A* (*Cyp2a*) genes. The *in vitro* transcription assay does not reflect the *in vivo* sex-specific expressions of these

genes either (Matsunaga *et al.*, 1990; Lindberg *et al.*, 1989). The possible explanations are :1) The method of extract preparation is not satisfactory enough, and some necessary factors such as hormone receptors may be lost or deactivated during the preparation; 2) Different regulation pathways such as post-transcriptional level regulation exist in rat or mice which can not be picked up in *in vitro* assays; and 3) The different tissue-specific control elements which can enhance or suppress each other are not included in the constructs used for the *in vitro* experiments.

3.2 Purpose of This Study

The increasing body of information about the structure, evolution and regulation of P450s makes it obvious that the enzyme activities of P450s are controlled by a variety of regulating mechanisms, many of which are species-specific and tissue-specific. The studies with human P450s clearly demonstrate that drug metabolism and the regulation of the drug-metabolising enzymes gene expression are often different in man compared to these in experimental animals. Three basic differences are identified. First, even highly structurally related P450s in rodents and humans, especially those in the *CYP2* family, may have variable catalytic activity (Wrighton *et al.*, 1992). For example, a well studied case is the metabolism of coumarin, which is a product of certain plants and fungi. Coumarin is metabolised in humans (Raunio *et al.*, 1988) and in mice (Lindberg and Negishi, 1989b) by CYP2A-mediated hydroxylation at the 7 position. The coumarin 7-hydroxylase activity is higher in human liver than that in mice (Lindberg and Negishi, 1989b; Raunio *et al.*, 1988). In rat, however, the coumarin 7-hydroxylase is very low, and the amount of excreted 7-hydroxycoumarin is less than 1% of the dose administered (Pearce *et al.*, 1992). Second, the expression of related P450 genes in varied species are regulated differently. A case of this is the sexual dimorphism which is observed in the metabolism of some compounds by rats and mice, but is not observed in humans (Wrighton and Stevens, 1992). Third, through gene duplication, species-specific *CYP* genes have evolved (Nebert, 1987). For example, rat *CYP2A3* is

found to be specifically expressed in lung tissue, but no similar gene is found existing in mice and humans. These differences among rodent species and man have made it difficult to extrapolate the conclusions drawn from the P450-mediated metabolism studies performed in rodent to human. Therefore, it is becoming increasingly important to characterise P450s and genes in the human *CYP2A* subfamily.

The human *CYP2A* subfamily is composed, at least, of two members, designated *CYP2A6* and *CYP2A7* (Miles *et al.*, 1989; Yamano *et al.*, 1990). Both genes are expressed in human liver and share 96% similarity of nucleotide sequences and 94% similarity of amino acid sequences. Several lines of evidence suggest that enzymes of the CYP2A subfamily play a role in the metabolic activation of promutagens, such as smoking related nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), N-nitrosodiethylamine (NDEA) (Yamazaki *et al.*, 1992; Crespi *et al.*, 1990), benzo[a]pyrene and the hepato-carcinogen aflatoxin B1 (Aoyama *et al.*, 1991; Crespi *et al.*, 1991).

This project aims at characterising the function and regulation of the human *CYP2A* gene. The strategy towards this aim includes:

a) Isolation of human *CYP2A* genes from human genomic libraries.

b) Analysis of the gene structure, including intron-exon junctions and mutations in the gene.

c) Characterisation of promoter/enhancer elements in *CYP2A* genes by DNA sequencing and transfection assay.

3.3 Results

3.3.1 Isolation of the clones containing human CYP2A genes

Screening of human genomic DNA libraries in cosmid and in bacteriophage EMBL3 vectors was carried out as described in "Materials and Methods" (Chapter 2) with a 0.7 kb *CYP2A6* cDNA fragment used as a probe. In order to isolate a complete gene, a human genomic cosmid library (Kioussis *et al.*, 1987) was screened. One clone, designated CoIIA, was isolated, and then purified by spreading the cosmid clone on an agar plate and repeating the colony hybridisation (Figure 3.5A). Restriction enzyme digestion and Southern hybridisation analyses using three different *CYP2A6* cDNA probes, a 150 bp *Sal*I-*Pst*I fragment containing part of exon 1 (Miles *et al.*, 1989a), a 400 bp and a 300 bp PCR fragment containing exon 6 to 7, and exon 8 to 9, respectively, suggested that the CoIIA contained an entire coding region and, at least, a 3 kb 5'-flanking fragment of a gene in the *CYP2A* subfamily (Figure 3.5B). Only one *CYP2A* gene appeared to be contained within the 35 kb insert. A schematic gene structure is shown in Figure 3.6.

3.3.2 Gene structure

(A) Comparisons of CoIIA sequence with those of other *CYP2A* genes

An 8 kb (pCoIIA, H8) and a 2.8 kb (pCoIIA, H2.8) *Hind*III fragment, and a 4.8 kb (pCoIIA, E/H4.8) *Hind*III/*Eco*RI fragment were subcloned into the vector pUC18, respectively. The restriction enzyme sites in these clones (shown in Figure 3.6) were used to generate different subclones for sequencing and for further analysis. In addition, oligonucleotides based on the intron/exon organisation of *CYP2A* subfamily genes were used to sequence all exons and parts of the introns. The exons were sequenced completely and the introns were partially sequenced. Comparison of the

Figure 3.5, A: The isolated cosmid clone CoIIA by repeating the colony hybridisation with a *CYP2A6* cDNA probe. The arrow points to a colony hybridised with the probe.

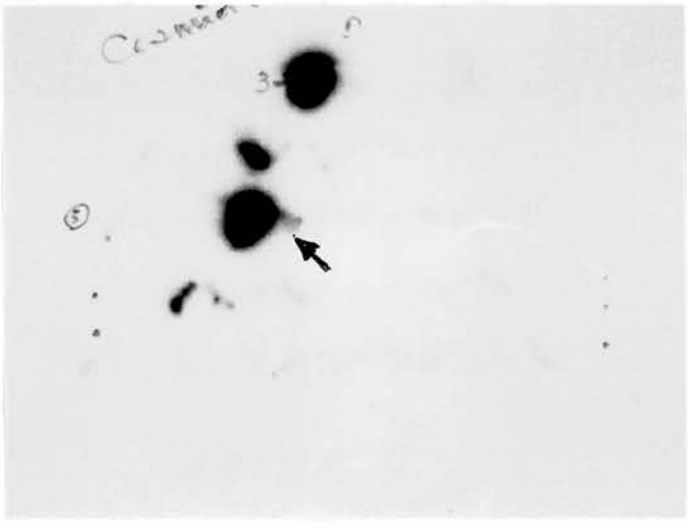


Figure 3.5, B: Southern blot analyses of the CoIIA DNA. The DNA was digested with *EcoR*I (E), *Hind*III (H) and double digested with *EcoRI*/*Hind*III (E/H). The blots were hybridised with random labelled *CYP2A6* cDNA probes; A, with a probe of exon 8 to 9; B, with a probe of exons 6 to 7 and C, with a probe of part exon 1.

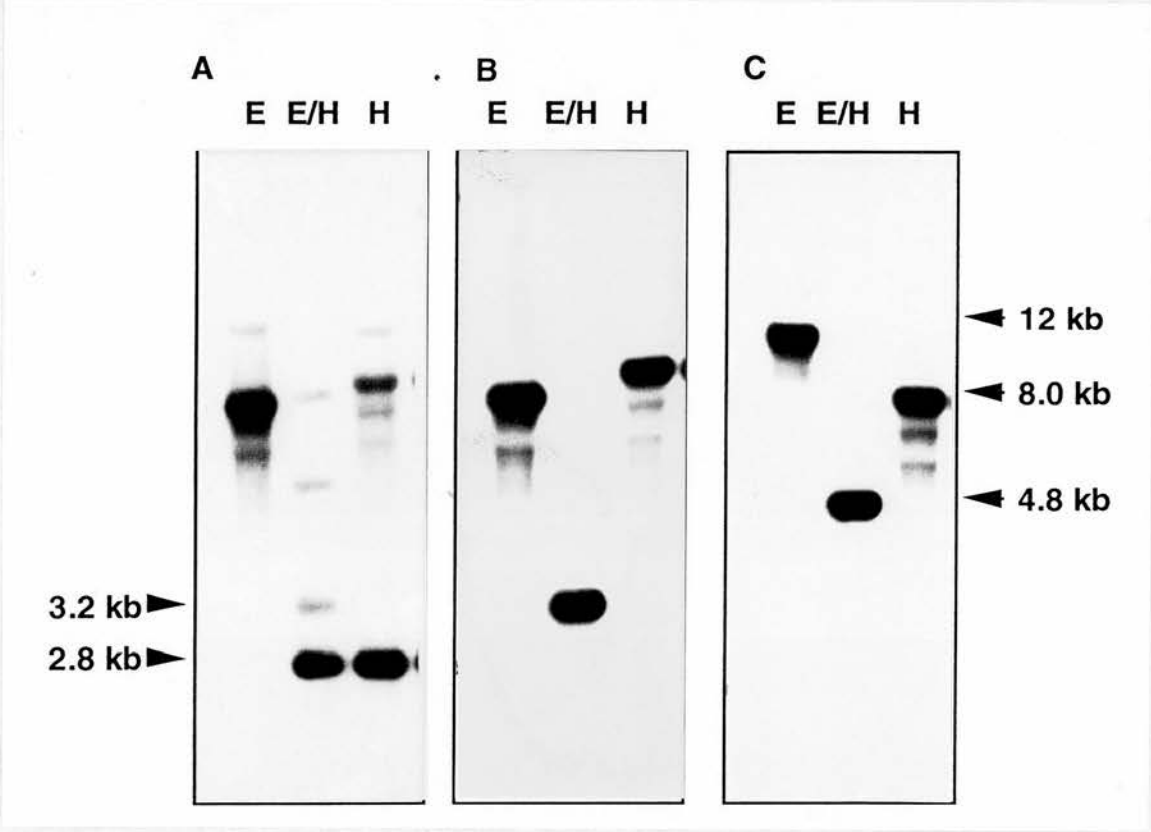
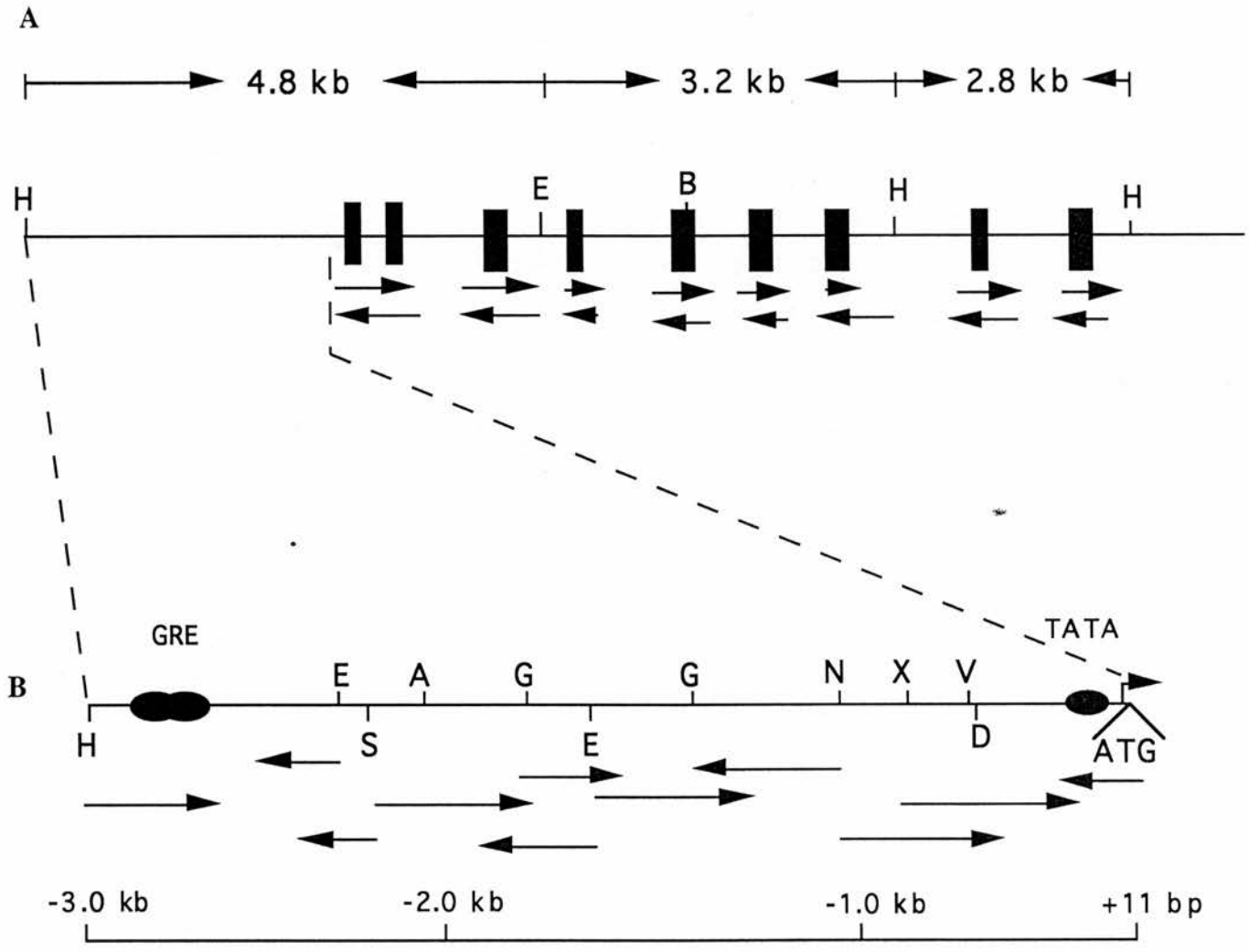


Figure 3.6 Schematic representation of the human *CYP2A7A* gene. The structure of the gene was elucidated by DNA sequencing and restriction analysis. The following restriction sites are shown *BamH*I (B); *EcoR*I (E); *Apa*I (A); *Bgl*II (G); *Hind*III (H); *BstX*I (X); *EcoR*V (R); *Nco*I (N); *Nde*I (D); *Sst*I (S) and *Ava*I (V). **A**: Exons are represented by solid boxes, and arrows under the exons indicate the sequencing direction from both ends for at least twice. The connecting line represents introns and the 5'-flanking region. **B**: The 5'-region of *CYP2A7A* gene. Arrows indicate the sequencing direction. GRE, Glucocorticoid responsive element; TATA, TATA box.

Figure 3.6


A
4.8 kb
3.2 kb
2.8 kb
H
E
B
H
H
B
GRE
TATA
E
A
G
G
N
X
V
H
S
E
D
ATG
-3.0 kb
-2.0 kb
-1.0 kb
+11 bp

exon sequences in CoIIA with the published sequence of the human *CYP2A6* and *CYP2A7* cDNAs (Yamano *et al.*, 1990) revealed that there were 54 bp (3.6%), and 34 deduced amino acid (6.9%) differences between CoIIA and CYP2A6. There were 13 base changes (0.9%) resulting in 5 amino acid (1 %) differences between CoIIA and *CYP2A7* (Table 3.4).

It has been shown that there are two hypervariable regions between CYP2A6 and CYP2A7: Six amino acid differences between residues 53 to 64, and five differences between residues 158 and 164 (Yamano *et al.*, 1990). Comparison of the deduced amino acid sequences of CoIIA with that of *CYP2A6* showed this was also the case here, but seven amino acid differences were found in the region between residues 53 to 64, and only four amino acid differences between amino acids 158 to 164. In addition, the proteins encoded by CoIIA and by *CYP2A7* showed a common difference to CYP2A6, namely, two base pair changes (GTA to GCG) in the codon 117, resulting in a substitution of alanine for valine. These data indicated that proteins encoded by *CYP2A7* and CoIIA could be alleles. The gene in CoIIA was termed *CYP2A7A*. Sequences of the exons, part of the introns and 5'-flanking region of the *CYP2A7A* are shown in Figure 3.7.

(B) Structure of *CYP2A7A*

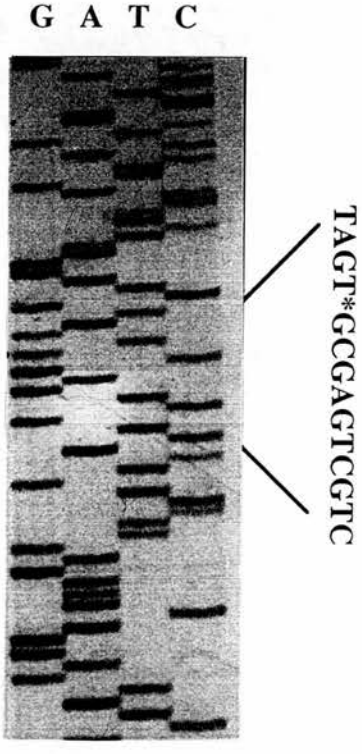
The intron-exon organisation of the *CYP2A7A* gene is shown in Table 3.5. The intron-exon junctions were initially predicted by comparing CoIIA sequence with that of a murine *Cyp2a* gene (Lindberg *et al.*, 1989a) and confirmed by sequencing and by comparing with the *CYP2A7* cDNA sequence. The approximate lengths of introns, except intron 1 which was completely sequenced, were established by PCR amplification followed by agarose gel electrophoresis and comparison to a 1 kb DNA ladder marker. Similar to the mouse genes in the *Cyp2a* subfamily, human *CYP2A7A* was 8 kb in size, and contained nine exons, interrupted by eight introns ranging in size

Table 3.4 Differences of nucleotide and deduced amino acid sequences between human *CYP2A7* and *CYP2A7A*. (**B**) and (**C**) show encoding sequences of the *CYP2A7A* . Asterisks indicate the nucleotide differences at the positions 610 (**B**) and 1423 (**C**).

	CYP2A7	*CYP2A7A*
488-489	**G-C (Ser^{163})**	**C-G (Thr^{163})**
610	**A (Met^{204})**	**G (Val^{204})**
821	**A (His^{274})**	**G (Arg^{274})**
909	**G**	**C**
915	**C**	**G**
1039	**A (Thr^{347})**	**G (Ala^{347})**
1119	**A**	**G**
1122	**C**	**T**
1209	**T**	**C**
1224	**C**	**T**
1245	**C**	**T**
1423	**T (Ser^{475})**	**C (Pro^{475})**

Continued table 3.4, B and C.

B



C

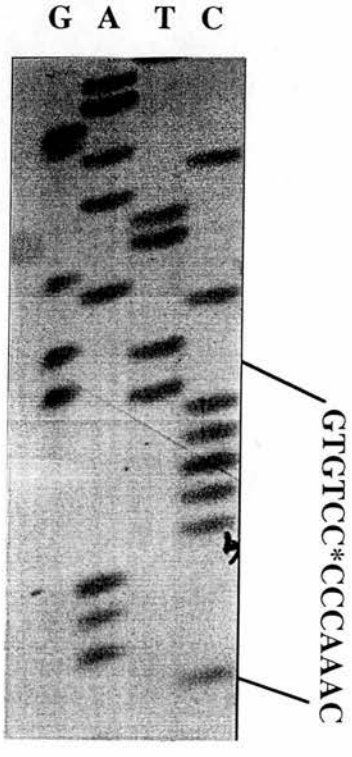


Table 3.5 Exon-intron organisation of the *CYP2A7A* gene

Exon sizes are given in base pairs (bp). Approximate intron sizes, except intron 1, are given in kilobases (kb). The exon sequences are designated by upper case letters; intron sequences by lower case letters. Splice donor and acceptor sites are shown in bold lower case letters.

Exon	Size (bp)	5' splice site	Intron	3'-splice site	Amino acid interrupted
1	180	TGAAG **gt**gtc	0.273 bp	act**ag** TTCAG	Lys^{60} / Ile^{61}
2	163	CTATG **gt**gag	0.9 kb	ccc**ag** GCGTG	Gly^{115}
3	150	GCACG **gt**gag	0.3 kb	cgc**ag** GCGCC	Gly^{165}
4	161	GGCAG **gt**aac	1.1 kb	ccc**ag** CTCTA	Gln^{218} / Leu^{219}
5	177	AGGAG **gt**aca	0.7 kb	tac**ag** GAGGA	Glu^{277} /Glu^{278}
6	142	GGAGG **gt**aag	0.5 kb	aca**ag** CCCAG	Ala^{325}
7	188	CTAAG **gt**gct	1.0 kb	acc**ag** GGCAC	Lys^{387} / Gly^{388}
8	142	CATCG **gt**aag	0.7 kb	ctc**ag** GAAAG	Gly^{435}
9	179				

between 273 bp and 1.1 kb. All the intron-exon junctions followed the GT-AG donor/acceptor rule.

Primer extension analysis was employed to identify the transcription initiation site using a 22-mer synthetic oligonucleotide complementary to the sequence 73 to 51 bp downstream from the first codon as the primer. Multiple extended fragments were identified, corresponding to initiation at various sites between 14 to 87 bp upstream of the initial methionine (Figure 3.8). Several smaller fragments (for example, 62 and 65 bp) were likely to reflect incomplete strand synthesis by the reverse transcriptase due to local secondary structure. Due to homology of exon 1 among *CYP2A* genes, the primer used in the extension analysis was not specific to *CYP2A7A*. Therefore, the transcription initiation site of *CYP2A7A* can not be clearly defined by the result of the primer extension analysis.

3.3.3 Alleles of CYP2A7

In addition to *CYP2A7A*, a second clone, LIIA, was isolated from a human EMBL3 genomic library by screening a total of 1.0 X 10^6 pfu (plaque forming unit) with a *CYP2A6* cDNA probe. Two more rounds of screening were carried out to purify the positive phage away from contaminating phage. In order to compare this clone with *CYP2A7A*, the DNAs of LIIA and *CYP2A7A* were separately digested with different restriction enzymes, and hybridised to a 5'-end specific probe (a 150 bp *Sal*I/*Pst*I fragment containing partial exon 1 of *CYP2A6*) and a 3'-end specific probe (a 900 bp *BamH*I/*Eco*RI fragments containing exon 5 to 9). The results showed that LIIA hybridised with the 3'-end probe, but not with the 5'-end probe (Figure 3.9A), suggesting that it was a member of the *CYP2A* subfamily gene missing exon 1. The restriction maps, using *Hind*III or *Eco*RI and double digestion with *Eco*RI/*Hind*III, indicated several differences between the gene in LIIA and *CYP2A7A*. For example, *Hind*III digestion generated 2.8 kb and an 8.0 kb fragment for the *CYP2A7A*; a 2.8 kb

Figure 3.7 Nucleotide sequences of human *CYP2A7A* gene. The genomic clone CoIIA was digested with restriction enzymes and the fragments were subcloned into vector pUC18/19. The subcloned fragments containing the exons were sequenced from both ends. Exonic sequences are denoted by bold letters, sequences of the introns and the 5' flanking region by plain letters.

```
-1263  CTGCCTCTGTGCCTAACACTGGAGTTACCCCAATCTCTTCTGCCACTGTTCCCTCTGCCT
-1203  GCTAATAGTAGTAGCCCCTGACAAAGCAGGAATCGCCCTTAAAGGAGACTTAACTCACCC
-1143  TCAAATGTGATCTTCTCTTCCCAACACTCCCTTTCCACTGGCAGAAAACCAAATCCAGAA
-1083  AGGGAACTAAGTACACAGAGCAGGAGAGATGGGAGTTCAGGGCCACTCACACATGTCCCC
-1023  TGCCCACTGTCTGTTTTCTGTCCTCTGTAGATCTTTATATAAAATGAGAAACATAAACAA
 -963  CAATCATAATATTAATAGGGATGATACAATCAATCTAGTGGGTTTCCCTGAGGATCTGGG
 -903  TTGGAAACCAGCGGGACAACCCTTGGGACACTTGAGATTTTCCACCATTTGGGGTTGTTA
 -843  TTACTATTTCTCCAGACCTAGCCAATCACTCTGCCAACTGCTAACTCCAGGTACTCTTCT
 -783  CCAGGTGTGGGGAAAGTTCCCCTGAAATATGGCCCTGGTCTTCCACCCCTTTCCCAACCA
 -723  GAGATGGGCAGTGTAGGTTCCATGGCACCATCCTGCCTCACTCTGAGGTTCCAGTGAGAT
 -663  TCTGGGCATCAAGAGATAGCTCTGGGCAAAAGCAAAATCAAGTCAGCCCCTGGACCCAGT
 -603  GCTGGGCTGCTGGGCTTTCTGGGAGCACCTGCTGGCTTGCTACACACTCCTCCTCCCAGA
 -543  AACTCCACACCCACAGCCCTGGGTCTTCCTAGCCCCGAGACTTTCAAGTCCATATGCCTG
 -483  GAGTCCCCCCTCCTGAGACCCTTAACCCTGCATCCTCCGCAACAGAAGACCCCCAGATGC
 -423  ACAGCACACTTCATTACTAATAAACCAGACTTGATCTTCTCCTGATGCCCAAATCCAGCA
 -363  AGCTTTGGGGTGCATTCTCACTCTCAGACCCCAAATCCAAAGTCCCAAGTGCTCCCCTAG
 -303  TATCCAAATATTCCAAACTCTTCAGTTCTACAGTTTATCGGTTGCCCCCTCCTAAATCCA
 -243  CAGCCTGCGGCACCCCTCCTGAAGTACCACAGATTTAGTCTGGAGGCCCCTCTCTGTTCA
 -183  GCTGCCCTGGGTCCCTTATCCTCCCTTGCTGGCTGTGTCCCAAGCTAGGTGGCATTCATG
 -123  GTGGGCGTGTAGTTGGGAGGTGAAATAAGGTGATTATGTAATTAGCCAAAGTCCATCCCT
  -63  CTTTTTCAGGCAGTATAAAGGCAAACCACCCCACCCATCACCATCTGTCATCTCACTACC
   -3  ACC
```
ATGCTGGCCTCAGGGCTGCTTCTGGTGGCCTTGCTGGCCTGCCTGACTGTGATGGTC

Exon 1
TTGATGTCTGTCTGGCAGCAGAGGAAGAGCAGGGGGAAGCTGCCTCCGGGACCCACCCCA
CTGCCCTTCATTGGAAACTACCTCCAGCTGAACACAGAGCACATATGTGACTCCATCATG
AAGGTGTCCCAAGGCAGGAAGATGGGTGGCACAGGGTGGGGGCTGCCTAGTTGGCTGGGG
CTTTGTGGCAGGGGATTGACCAGTGTGGACCAGAGTCTTAGGAAAGGGAGTTTTGGAGTT
TCAGCATCCGGGTCCTAGCCAGGAAAGACAGGATCTTGGGATGTCCAGCTCCCTGACTGT
GAGAACCTGGGGGTGAAGGATCCCAGTACTTGACATCTCGGTGCTGGGCCCCATTCAGAG

Exon 2
TGGGGGCTGCTCCCTCTAACCACTCCCACCCGCCTCCATCAG
TTCAGTGAGTGCTATGGC
CCCGTGTTCACCATTCACTTGGGGCCCCGGCGGGTCGTGGTGCTGTGTGGACATGATGCC
GTCAGGGAGGCTCTGGTGGACCAGGCTGAGGAGTTCAGCGGGCGAGGCGAGCAAGCCACC
TTCGACTGGGTCTTCAAAGGCTATGGTGAGGGGGTGCCCAAGATGGGGAAGGTGGCCAGG
CGGATATGATGGTCTCAGTGTTGCCAGCCTTCTCCCTGACTCTCCTGCCCACTGGAGGCT
ATGGCAGAACCCCGTCTGTCTTATCTCCCATCTCCTTCACACTGTGGCCTCTCATGTGTA
TCCTCACCTGTCTAGCGCTCTGTCGTATTCATCTAGCTCTCTCTGACCACTCTATCTCTC
TACATGAG---------- 0.7 kb ----------CCCAG

Exon 3
GCGTGGCGTTCAGCAACGGGGAGCGCGCCAAGCAGCTCCTGCGCTTTGCCATCGCCACCC
TGAGGGACTTCGGGGTGGGCAAGCGAGGCATCGAGGAGCGCATCCAGGAGGAGTCGGGCT
TCCTCATCGAGGCCATCCGGAGCACGCACGGTGAGTAAGGTTCCCCGCGTGGCCAGGAGA
AGGAAAACACCCAGGACGAGGAACCCGCGCGCGTTCTGCCTGCGGATGGGGACTAGGTGG
GGAAAGGCGCCCGCACTTCCAGCCCTGGAGTCTGGCGCTGTGAATTTGGCTCAACAAGGC
CCTGCCTCCTGGAATTC-------- 50 bp ----------
CTGATTCTCCTCAGACCTCTGAGTTGACTCTCTCCCCAACCCGCTTCTCCCTCCACACCC

Exon 4
GCAG
GCGCCAATATCGATCCCACCTTCTTCCTGAGCCGCACAGTCTCCAATGTCATCAGC
TCCATTGTCTTTGGGGACCGCTTTGACTATGAGGACAAAGAGTTCCTGTCACTGCTGAGC
GTGATGCTAGGAATCTTCCAGTTCACGTCAACCTCCACGGGGCAGGTAACTGGCTGCTGC
CGGCCCGTGACGCCCCTACCAVAACCTGCCAACTGCTCCCCTACCTGGAGACAGGTGCCC
CAAACTCCCACCCCGCTCCAGACAGTGTCCCCTCAAAATCAGCCCCCGATATTGGACAAC
```
```

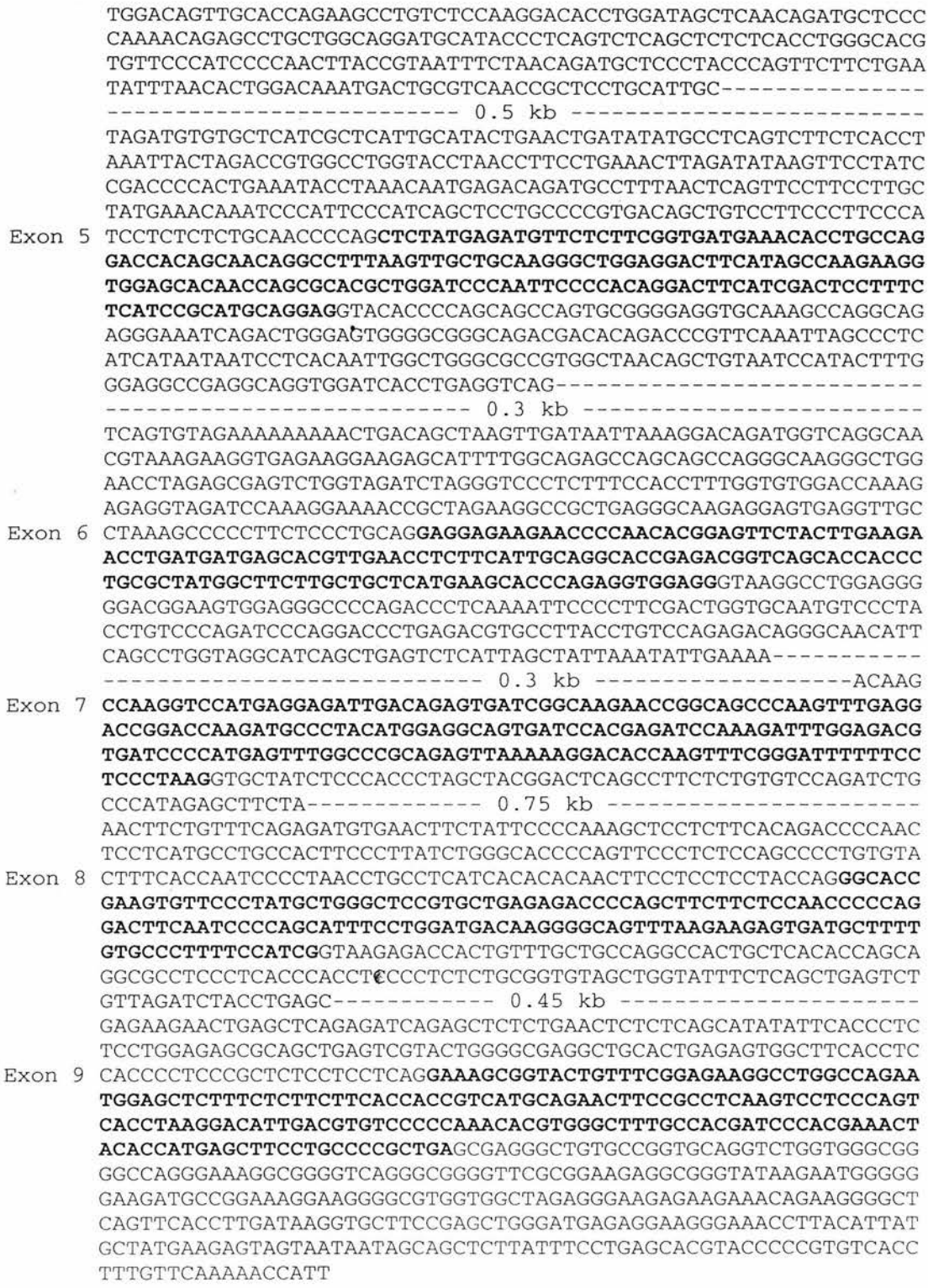

TGGACAGTTGCACCAGAAGCCTGTCTCCAAGGACACCTGGATAGCTCAACAGATGCTCCC
CAAAACAGAGCCTGCTGGCAGGATGCATACCCTCAGTCTCAGCTCTCTCACCTGGGCACG
TGTTCCCATCCCCAACTTACCGTAATTTCTAACAGATGCTCCCTACCCAGTTCTTCTGAA
TATTTAACACTGGACAAATGACTGCGTCAACCGCTCCTGCATTGC----------
---------- 0.5 kb ----------
TAGATGTGTGCTCATCGCTCATTGCATACTGAACTGATATATGCCTCAGTCTTCTCACCT
AAATTACTAGACCGTGGCCTGGTACCTAACCTTCCTGAAACTTAGATATAAGTTCCTATC
CGACCCCACTGAAATACCTAAACAATGAGACAGATGCCTTTAACTCAGTTCCTTCCTTGC
TATGAAACAAATCCCATTCCCATCAGCTCCTGCCCCGTGACAGCTGTCCTTCCCTTCCCA
Exon 5 TCCTCTCTCTGCAACCCCAG**CTCTATGAGATGTTCTCTTCGGTGATGAAACACCTGCCAG**
GACCACAGCAACAGGCCTTTAAGTTGCTGCAAGGGCTGGAGGACTTCATAGCCAAGAAGG
TGGAGCACAACCAGCGCACGCTGGATCCCAATTCCCCACAGGACTTCATCGACTCCTTTC
TCATCCGCATGCAGGAGGTACACCCCAGCAGCCAGTGCGGGGAGGTGCAAAGCCAGGCAG
AGGGAAATCAGACTGGGAGTGGGGCGGGCAGACGACACAGACCCGTTCAAATTAGCCCTC
ATCATAATAATCCTCACAATTGGCTGGGCGCCGTGGCTAACAGCTGTAATCCATACTTTG
GGAGGCCGAGGCAGGTGGATCACCTGAGGTCAG----------
---------- 0.3 kb ----------
TCAGTGTAGAAAAAAAAACTGACAGCTAAGTTGATAATTAAAGGACAGATGGTCAGGCAA
CGTAAAGAAGGTGAGAAGGAAGAGCATTTTGGCAGAGCCAGCAGCCAGGGCAAGGGCTGG
AACCTAGAGCGAGTCTGGTAGATCTAGGGTCCCTCTTTCCACCTTTGGTGTGGACCAAAG
AGAGGTAGATCCAAAGGAAAACCGCTAGAAGGCCGCTGAGGGCAAGAGGAGTGAGGTTGC
Exon 6 CTAAAGCCCCCTTCTCCCTGCAG**GAGGAGAAGAACCCCAACACGGAGTTCTACTTGAAGA**
ACCTGATGATGAGCACGTTGAACCTCTTCATTGCAGGCACCGAGACGGTCAGCACCACCC
TGCGCTATGGCTTCTTGCTGCTCATGAAGCACCCAGAGGTGGAGGGTAAGGCCTGGAGGG
GGACGGAAGTGGAGGGCCCCAGACCCTCAAAATTCCCCTTCGACTGGTGCAATGTCCCTA
CCTGTCCCAGATCCCAGGACCCTGAGACGTGCCTTACCTGTCCAGAGACAGGGCAACATT
CAGCCTGGTAGGCATCAGCTGAGTCTCATTAGCTATTAAATATTGAAAA----------
---------- 0.3 kb ----------ACAAG
Exon 7 **CCAAGGTCCATGAGGAGATTGACAGAGTGATCGGCAAGAACCGGCAGCCCAAGTTTGAGG**
ACCGGACCAAGATGCCCTACATGGAGGCAGTGATCCACGAGATCCAAAGATTTGGAGACG
TGATCCCCATGAGTTTGGCCCGCAGAGTTAAAAAGGACACCAAGTTTCGGGATTTTTTCC
TCCCTAAGGTGCTATCTCCCACCCTAGCTACGGACTCAGCCTTCTCTGTGTCCAGATCTG
CCCATAGAGCTTCTA---------- 0.75 kb ----------
AACTTCTGTTTCAGAGATGTGAACTTCTATTCCCCAAAGCTCCTCTTCACAGACCCCAAC
TCCTCATGCCTGCCACTTCCCTTATCTGGGCACCCCAGTTCCCTCTCCAGCCCCTGTGTA
Exon 8 CTTTCACCAATCCCCTAACCTGCCTCATCACACACAACTTCCTCCTCCTACCAG**GGCACC**
GAAGTGTTCCCTATGCTGGGCTCCGTGCTGAGAGACCCCAGCTTCTTCTCCAACCCCCAG
GACTTCAATCCCCAGCATTTCCTGGATGACAAGGGGCAGTTTAAGAAGAGTGATGCTTTT
GTGCCCTTTTCCATCGGTAAGAGACCACTGTTTGCTGCCAGGCCACTGCTCACACCAGCA
GGCGCCTCCCTCACCCACCTCCCCTCTCTGCGGTGTAGCTGGTATTTCTCAGCTGAGTCT
GTTAGATCTACCTGAGC---------- 0.45 kb ----------
GAGAAGAACTGAGCTCAGAGATCAGAGCTCTCTGAACTCTCTCAGCATATATTCACCCTC
TCCTGGAGAGCGCAGCTGAGTCGTACTGGGGCGAGGCTGCACTGAGAGTGGCTTCACCTC
Exon 9 CACCCCTCCCGCTCTCCTCCTCAG**GAAAGCGGTACTGTTTCGGAGAAGGCCTGGCCAGAA**
TGGAGCTCTTTCTCTTCTTCACCACCGTCATGCAGAACTTCCGCCTCAAGTCCTCCCAGT
CACCTAAGGACATTGACGTGTCCCCCAAACACGTGGGCTTTGCCACGATCCCACGAAACT
ACACCATGAGCTTCCTGCCCCGCTGAGCGAGGGCTGTGCCGGTGCAGGTCTGGTGGGCGG
GGCCAGGGAAAGGCGGGGTCAGGGCGGGGTTCGCGGAAGAGGCGGGTATAAGAATGGGGG
GAAGATGCCGGAAAGGAAGGGGCGTGGTGGCTAGAGGGAAGAGAAGAAACAGAAGGGGCT
CAGTTCACCTTGATAAGGTGCTTCCGAGCTGGGATGAGAGGAAGGGAAACCTTACATTAT
GCTATGAAGAGTAGTAATAATAGCAGCTCTTATTTCCTGAGCACGTACCCCCGTGTCACC
TTTGTTCAAAAACCATT

Figure 3.8

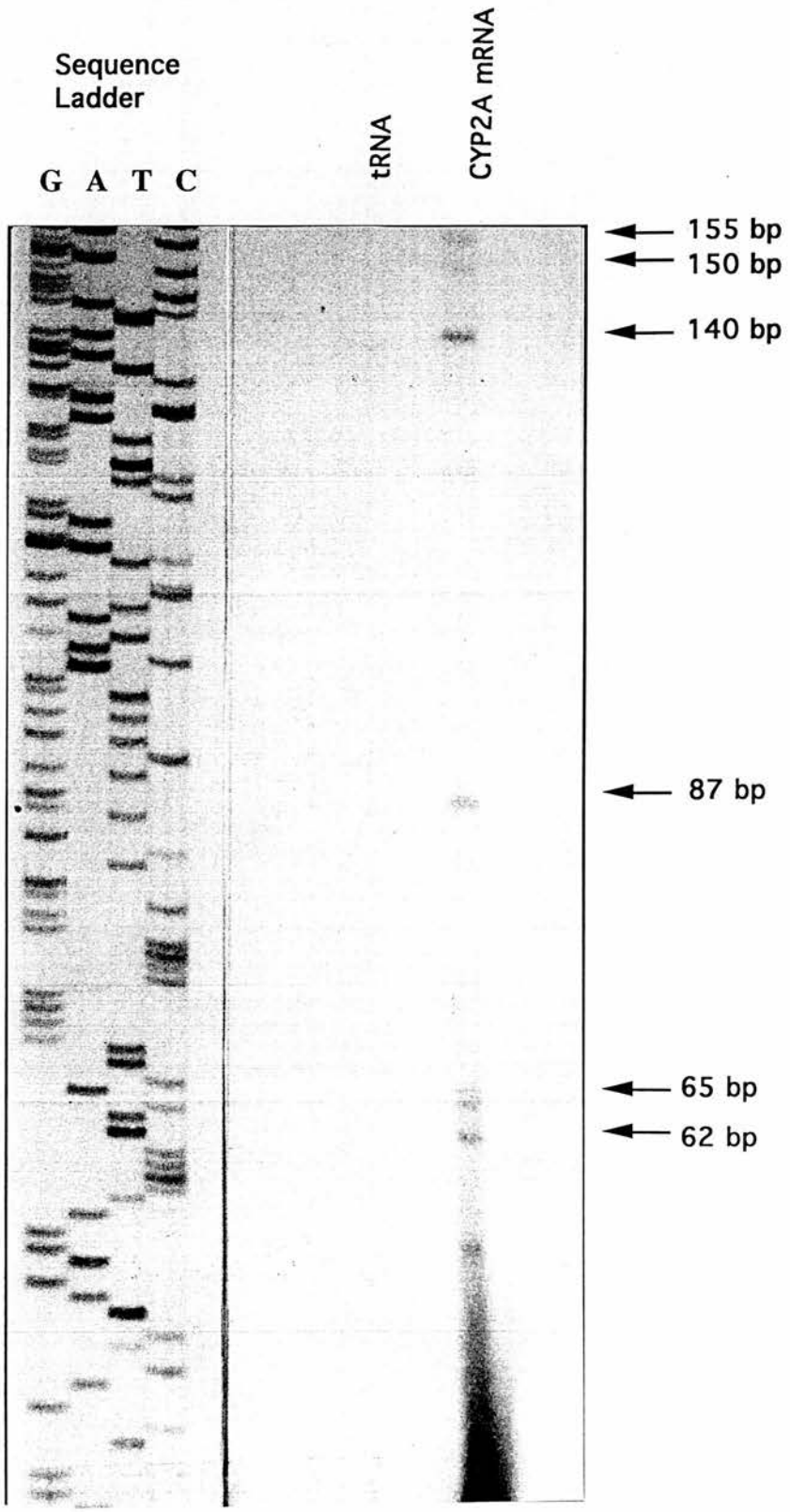

Sequence Ladder
G A T C
tRNA
CYP2A mRNA
155 bp
150 bp
140 bp
87 bp
65 bp
62 bp

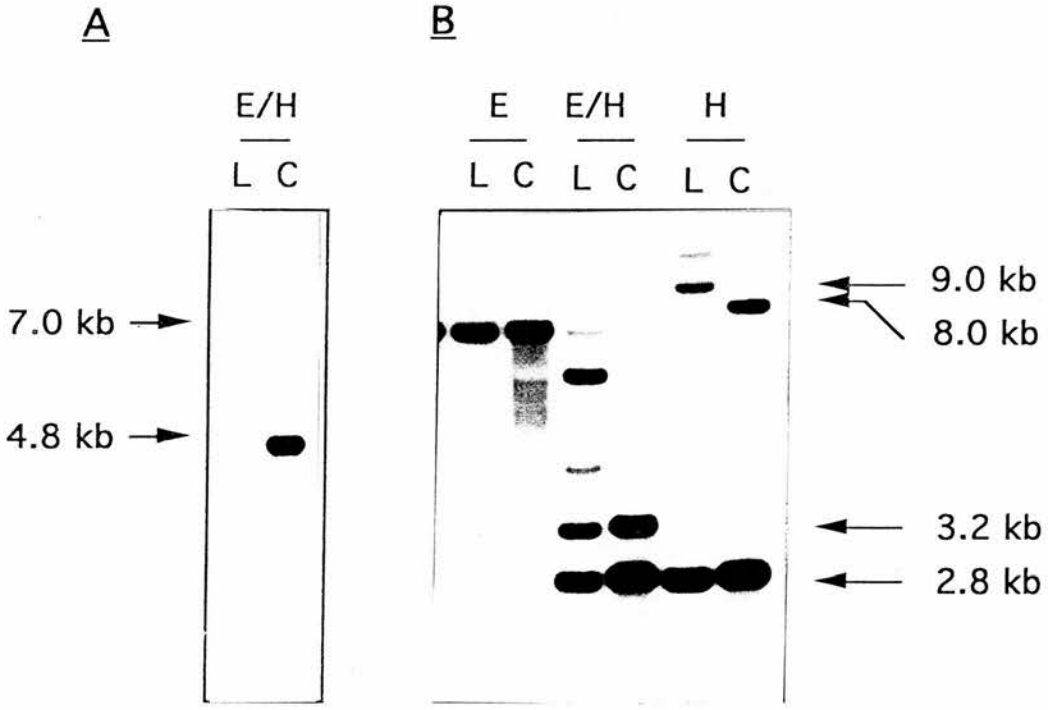
Figure 3.8 Primer extension analysis. The primer used in the analysis was a 22-mer oligonucleotide complementary to the sequence 73 to 51 bp downstream from the first codon. Total RNA (50 μg) from a human liver was used. Yeast tRNA was used as a negative control. A sequence ladder was used as a size standard. The arrows show the multiple extended fragments.

fragment and a 9 kb fragment for LIIA (Figure 3.9B). Only digested *CYP2A7A* DNA hybridised with the 5'-end probe (Figure 3.9A). These results showed that the LIIA contained a genomic DNA fragment which spanned from exon 2 to the 3' end of a *CYP2A* gene, and this gene was not identical with *CYP2A7A* .

The gene in LIIA was partially sequenced from exon 2 to exon 3. This analysis showed (Table 3.6) that there were several base pair differences between it and both *CYP2A6* and *CYP2A7A*, respectively. One of the sequence differences with *CYP2A7A* was only one silent base pair change (GTA to GTG), which did not result in amino acid change in the codon 117. The *CYP2A7A* had two base pair changes in this codon (GTA to GCG), resulting in a substitution of alanine for valine. At this position the amino acid encoded by LIIA was the same as *CYP2A6*. On the other hand, LIIA also contained a hypervariable region between the deduced amino acid residues 158 and 164, which was identical to the CYP2A7A. Based on these results it is suggested that the gene in clone LIIA is another allele of *CYP2A7*, designated *CYP2A7B*. A comparison of nucleotide and deduced amino acid differences among the *CYP2A7B*, *CYP2A7A*, *CYP2A7* and *CYP2A6* is shown in Table 3.6.

In order to distinguish between *CYP2A6* and the three alleles of *CYP2A7*, a PCR

Figure 3.9 Southern blot analysis of LIIA (L) and CoIIA (C) DNAs. The DNAs were digested with *EcoR*I (E); with *Hind*III (H), or with *EcoR*I and *Hind*III both (E/H). The blots were hybridised with random labelled 5'-end (**A**) and 3'-end probes (**B**), separately.



strategy was carried out using human genomic DNA as template, an oligonucleotide A (5'-AGGTGATTATGTAATTAGCC---3') complementary to the 5' flanking region of *CYP2A7A* and an oligonucleotide B (5'---TTCTGCCATAGCCTCCAGTG---3') complementary to the intron 2 of *CYP2A7A* and *CYP2A7B*. The expected PCR product was a fragment of 817 bp in size (Figure 3.10A). As a consequence of one base pair difference (G to C at position 141, +1 indicates the start of the open reading frame) in the exon 1 of both *CYP2A7* and the *CYP2A7A*, a *Pst*I site is absent from both genes. The PCR amplified *CYP2A6* can be digested with *Pst*I to two fragments of 582 bp and 235 bp, whereas the amplified alleles of *CYP2A7* was not digested. After hybridisation to the exon 2 specific oligonuceotide probe, the 582 bp fragment of *CYP2A6* and 817 bp undigested fragment of *CYP2A7* alleles could be observed. The result of this analysis showed that none of the PCR product was digested with *Pst*I, indicating that the primers used in the analysis were specific for alleles of *CYP2A7* (Figure 3.10B). In order to confirm whether the *Pst*I site existed in amplified *CYP2A6* cDNA, the full length *CYP2A* cDNAs were amplified with an oligonucleotide, complementary to the 5' end of *CYP2A* (5'-CATGCTGGCCTCAGGGCTGCTT-3') and an antisense oligonucleotide, complementary to the 3' end of *CYP2A* (5'-GCCTTAAGGCTTCCCCCCATTCTTATACC-3'). The amplified products were cloned into vector pUC18 and screened by *Pst*I digestion. The sequencing results indicated that the clone digested with *Pst*I contained the *CYP2A6* cDNA, whereas the clone not digested with *Pst*I was *CYP2A7* (data not shown).

It was then investigated whether *CYP2A7A* and *CYP2A7B* were actually alleles or represented different genes. In order to do this, a *Mbo*I restriction enzyme (restriction site, **GATC**) was used. One *Mbo*I restriction site existed at the intron 1/exon 2 boundary (a**g ATC**) in *CYP2A7B*, but the site was absent in *CYP2A7A* (a**g TTC**). The PCR with primers A and B, followed by *Mbo*I digestion generated a 346 bp fragment for *CYP2A7A* and a 263 bp fragment for *CYP2A7B* sequence. Both fragments were hybridised to the exon 2 specific oligonuceotide probe (Figure 3.10, A

Table 3.6. The differences of nucleotide and deduced amino acid (exon 2 to exon 3) among CYP2A6, CYP2A7, CYP2A7A and CYP2A7B.

	181	190	217	349-351	383	390-391	406	457	473	478	484	487-491
CYP2A6	A Ile[61]	C Arg[64]	T	GTA Val[117]	G Ala[128]	CT Ser[131]	C	G Ala[153]	C Asp[158]	C Ala[160]	G Gly[162]	CTGG Thr[163] Gly[164]
CYP2A7B	A Ile[61]	C Arg[64]	C	GTG Val[117]	G Ala[128]	CT Ser[131]	A	G Ala[153]	G Glu[158]	A Ile[160]	A Ser[162]	CGCA Thr[163] His[164]
CYP2A7A	T Thr[61]	T Thr[64]	T	GCG Ala[117]	T Leu[128]	TG Ala[131]	A	T Ser[153]	G Glu[158]	A Ile[160]	A Ser[162]	CGCA Thr[163] His[164]
CYP2A7	T Thr[61]	T Thr[64]	T	GCG Ala[117]	T Leu[128]	TG Ala[131]	A	T Ser[153]	G Glu[158]	A Ile[160]	A Ser[162]	GCCA Ser[163] His[164]

and C). The analysis of the PCR product from 18 different individuals showed eight were homozygous for the *CYP2A7B* (Figure 3.10, C, track 3), six were homozygous for the *CYP2A7A* (Figure 3.10, C, track 2, 4 and 5) and others were heterozygous for *CYP2A7A/CYP2A7B.* (Figure 3.10, C, track 1). This gives allele frequencies of approximately 44.4%, 33.3% and 22.2%, respectively. These data provided more evidence that the genes encoded by *CYP2A7A* and *CYP2A7B* were alleles and it was likely that both genes were alleles of *CYP2A7*. However, in the absence of the *CYP2A7* genomic sequence it was not certain that the amplification primers would actually amplify the *CYP2A7* gene.

3.3.4 The promoter elements in the 5'-flanking region of CYP2A7A gene

A 4.8 kb *Hind*III/*Eco*RI fragment of *CYP2A7A* was subcloned into pUC18 (pCoIIA, H/E4.8) and sequenced except for a narrow gap in the 5'-flanking region. This 4.8 kb DNA fragment contained 3.0 kb of the 5'-flanking region of the *CYP2A7A* gene (Figure 3.6B). A putative TATA box was located at 50 bp upstream from the first codon of methionine, but there was no CCAAT box. A computer analysis identified several potential transcription factor binding sites. A modified glucocorticoid-responsive element (GRE) (Jaiswal *et al.*, 1990), GGTAGGNNNTGTTCT, and a reverse orientated core sequence of GRE (TGTTCT) on the non-coding strand were found at -2.75 kb and -2.77 kb (Figure 3.6). In addition, within 0.5 kb 5' flanking region (Figure 3.11) a consensus sterol responsive element (SRE, 5'-CACCCCAC-3'), which is found in the promoters of several genes involved in steroid metabolism (Osborne *et al.*, 1988; Smith *et al.*, 1988), and a SRE-like element (7/8 bases of sequence identity of consensus SRE) were found at the positions of -37 bp and -232 bp. These sequences overlapped with three copies of another promoter motif, CACCC (Schüle *et al.*, 1988b; Walters *et al.*, 1992; Yu *et al.*, 1991). An element, 5'-

Figure 3.10

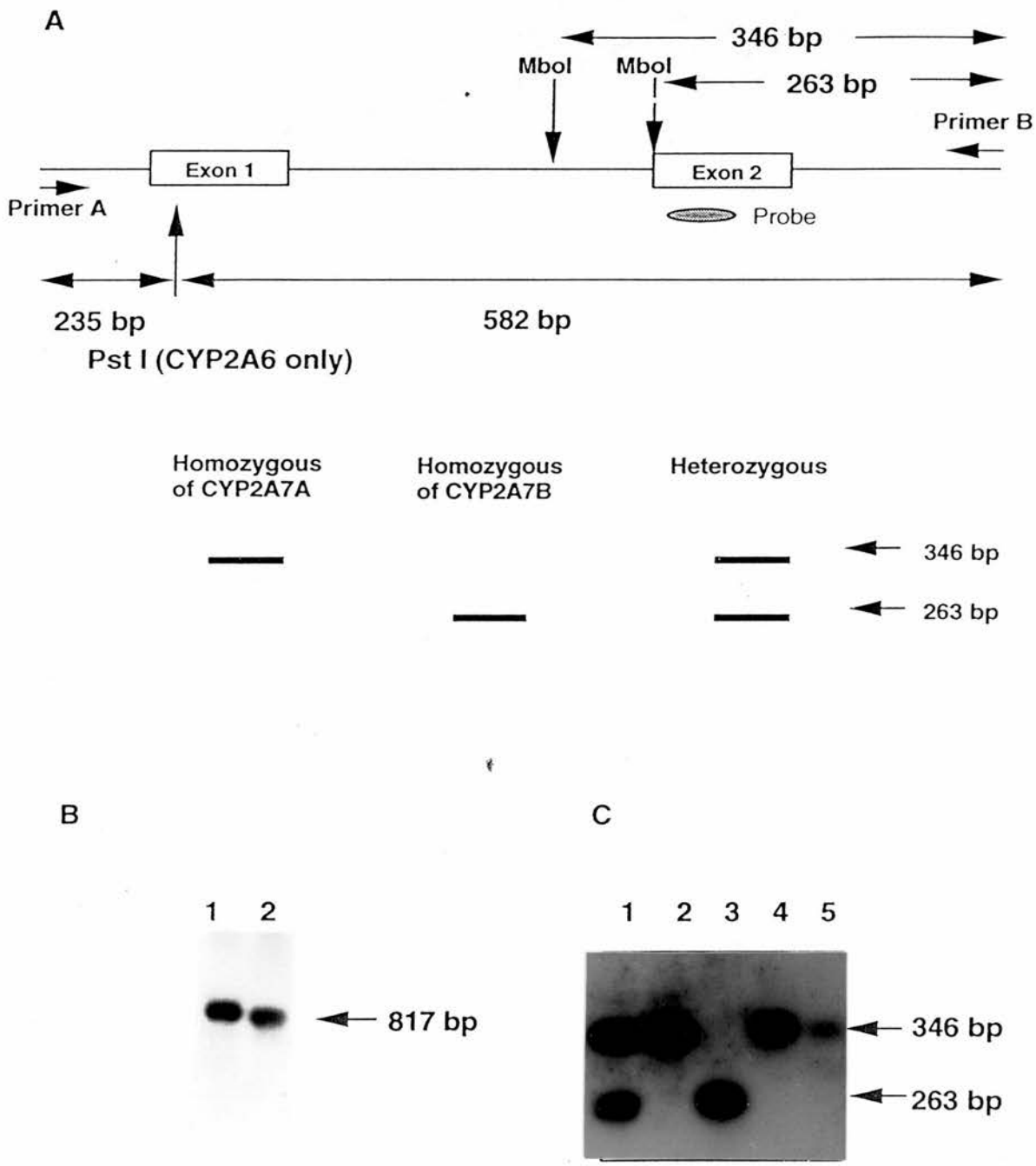
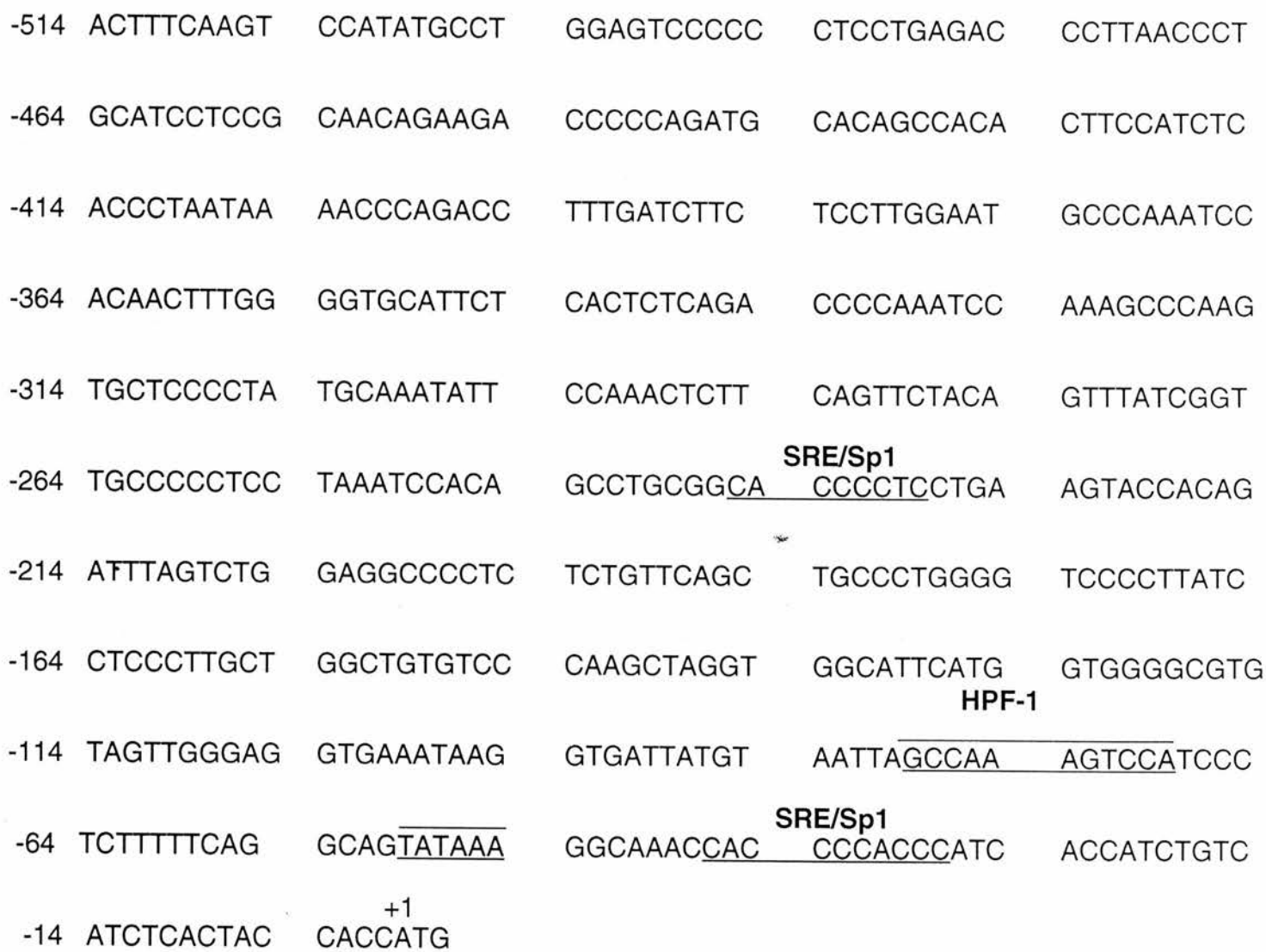

A
346 bp
MboI
MboI
263 bp
Primer B
Exon 1
Exon 2
Primer A
Probe
235 bp
582 bp
Pst I (CYP2A6 only)
Homozygous of CYP2A7A
Homozygous of CYP2A7B
Heterozygous
346 bp
263 bp
B
1 2
817 bp
C
1 2 3 4 5
346 bp
263 bp

Figure 3.10 Analyses of *CYP2A7* alleles by PCR. The PCR was carried out as described in Chapter 2. Genomic DNAs were prepared by Dr Zhong from human blood samples. **A.** Schematic diagram of the analysis procedure and the predicted banding patterns of amplified DNA following the digestion with *Mbo*I. **B**. The Southern hybridisation of *Pst*I digested PCR product with an exon 2 specific probe. 1, undigested DNA; 2, *Pst*I digested DNA. **C**. Southern hybridisation of PCR/*Mbo*I digested fragments. Homozygous *CYP2A7A* genotype is shown in tracks 2, 4 and 5; homozygous *CYP2A7B* genotype in track 3. Track 1 shows a heterozygote of *CYP2A7A/CYP2A7B*.

GCCAAAGTCCA-3', was found at position of -79, which has been reported (Venepally *et al.*, 1992) recently as a HepG2-specific P4502C factor-1 (HPF-1) binding site existing in the several *CYP2* gene subfamilies, such as *CYP2A*, *CYP2C* and *CYP2D*. It has been also reported that HPF-1 motif is a functional hepatic nuclear factor-4 (HNF-4) binding site (Chen *et al.*, 1994). In addition to these conserved elements, a modified TCDD-AH complex binding site (TTGCTGG) was also found in this region.

Figure 3.11 The nucleotide sequence of the 0.5 kb 5'-flanking region of *CYP2A7A* gene. The putative TATA box and HPF-1 binding site are boxed, the SRE/ CACCC (Sp1) motifs are underlined.

Figure 3.11


-514 ACTTTCAAGT CCATATGCCT GGAGTCCCCC CTCCTGAGAC CCTTAACCCT
-464 GCATCCTCCG CAACAGAAGA CCCCCAGATG CACAGCCACA CTTCCATCTC
-414 ACCCTAATAA AACCCAGACC TTTGATCTTC TCCTTGGAAT GCCCAAATCC
-364 ACAACTTTGG GGTGCATTCT CACTCTCAGA CCCCAAATCC AAAGCCCAAG
-314 TGCTCCCCTA TGCAAATATT CCAAACTCTT CAGTTCTACA GTTTATCGGT
SRE/Sp1
-264 TGCCCCCTCC TAAATCCACA GCCTGCGGCA CCCCTCCTGA AGTACCACAG
-214 ATTTAGTCTG GAGGCCCCTC TCTGTTCAGC TGCCCTGGGG TCCCCTTATC
-164 CTCCCTTGCT GGCTGTGTCC CAAGCTAGGT GGCATTCATG GTGGGGCGTG
HPF-1
-114 TAGTTGGGAG GTGAAATAAG GTGATTATGT AATTAGCCAA AGTCCATCCC
SRE/Sp1
-64 TCTTTTTCAG GCAGTATAAA GGCAAACCAC CCCACCCATC ACCATCTGTC
+1
-14 ATCTCACTAC CACCATG

3.3.5 Transcriptional regulation of CYP2A7A

(A) Construction of fused plasmids for *in vitro* transcription assays

PCR was carried out with the clone pCoIIA,H/E4.8 as the template DNA. The 5'-primer (5'-CCCAAGCTTGGCTGTGCTGAGGAAG-3', 3 kb upstream from the first codon, including a *Hind*III restriction site) and the 3' primer (5'-GGTAGTGAGATGACAGATGGT-3', starting from 2 bp upstream of first codon) were used. The 3 kb 5'-flanking region of the *CYP2A7A* fragment generated by this PCR was blunted using *Klenow* polymerase I, then digested with *Hind*III, and was cloned into the *Hind*III/blunted *Xba*I sites of the vector pCAT-basic containing CAT cDNA (Promega Ltd.) to form the construct pCAT2A7A5'-3.0. The preparation of 5' deleted constructs were described in Chapter 2.

(B) Promoter activity and xenobiotic induction

The conditions of transient transfection assay are described in Chapter 2. As shown in Figure 3.12A, the construct containing 3.0 kb 5'-flanking region of *CYP2A7A* exhibited only a low CAT activity. With the progressive deletion of the 3.0 kb 5'-flanking sequence to the position of -1.0 kb (pCAT2A7A5'-1.0), the transcriptional activity significantly increased and the maximal activity was observed with a fragment containing approximately 0.5 kb of the 5' end of the gene. To further define the promoter elements, the 0.5 kb 5' flanking region was analysed in more detail (Figure 3.12B). Deletion down to -95 bp (pCAT2A7A5'-D11) had no effect on the transcriptional activity. However, further deletion to -72 bp, which removed most of the HPF-1 element, resulted in a 80% decrease in activity, suggesting that the HPF-1 was a basal promoter element for the expression of the *CYP2A7A* gene in HepG2 cells.

To examine the potential inducing effects of xenobiotic on promoter activity of the *CYP2A7A*, chemicals, which have been found inducing *CYP2A* gene expressions in

animal experiments, were used. In my experiments, five chemicals, ß-naphthoflavone, phenobarbital, pyrazole, DEX and TCPOBOP, were used and the concentrations of these chemicals were chosen as described by Maurice *et al.* (1991), who treated the cultured human primary hepatocytes with ß-naphthoflavone, phenobarbital and pyrazole, and as described by Jaiswal *et al.* (1990), who treated rat hepatoma H4II cells with DEX. The concentration of TCPOBOP was used as described by Smith *et al.* (1993). In my experiments, cells transfected by pCAT2A7A5'-3.0 were grown in the presence of TCPOBOP (20 μM), phenobarbital (2 mM), DEX (0.1 μM), pyrazole (50 μM) and ß-naphthoflavone (10 μM) for 24-36 hr, and then the CAT activities in treated cells were assayed. The results showed that the treatment with DEX (52% increase; N=4; $P \leq 0.001$) and pyrazole (61% increase; N=3; $P \leq 0.001$) resulted in a slight increase in activity; phenobarbital treatment was inhibitory (52% decrease; N=4; $P \leq 0.001$) and TCPOBOP and ß-naphthoflavone produced no effect (Figure 3.13). However, the induction assay was carried out by using a single dosage of each chemicals, and different concentrations of each chemical might change the induction patterns. A full dose-response curve need to be carried out in future work to characterise the inducible expression of the human *CYP2A7A* gene in more details.

Figure 3.12 Constitutive expression of *CYP2A7A* gene in HepG2 cells. CAT activities were determined by liquid scintillation counting and normalised to ß-galactosidase activity. Values are expressed as the percentage of activity relative to the activity of pCAT2A7A5'-0.5. Each value is an average of three experiments. CAT activities measured by TLC are also shown below the maps of the deleted 5' flanking region. **A** shows the reporter plasmids constructed with the deleted 3 kb *CYP2A7A* promoter region, **B** represents the report plasmids constructed with the deleted 0.5 kb 5'flanking region in pCAT2A7A5'-0.5. **A** and **B** have different scales.

Fig. 3.12

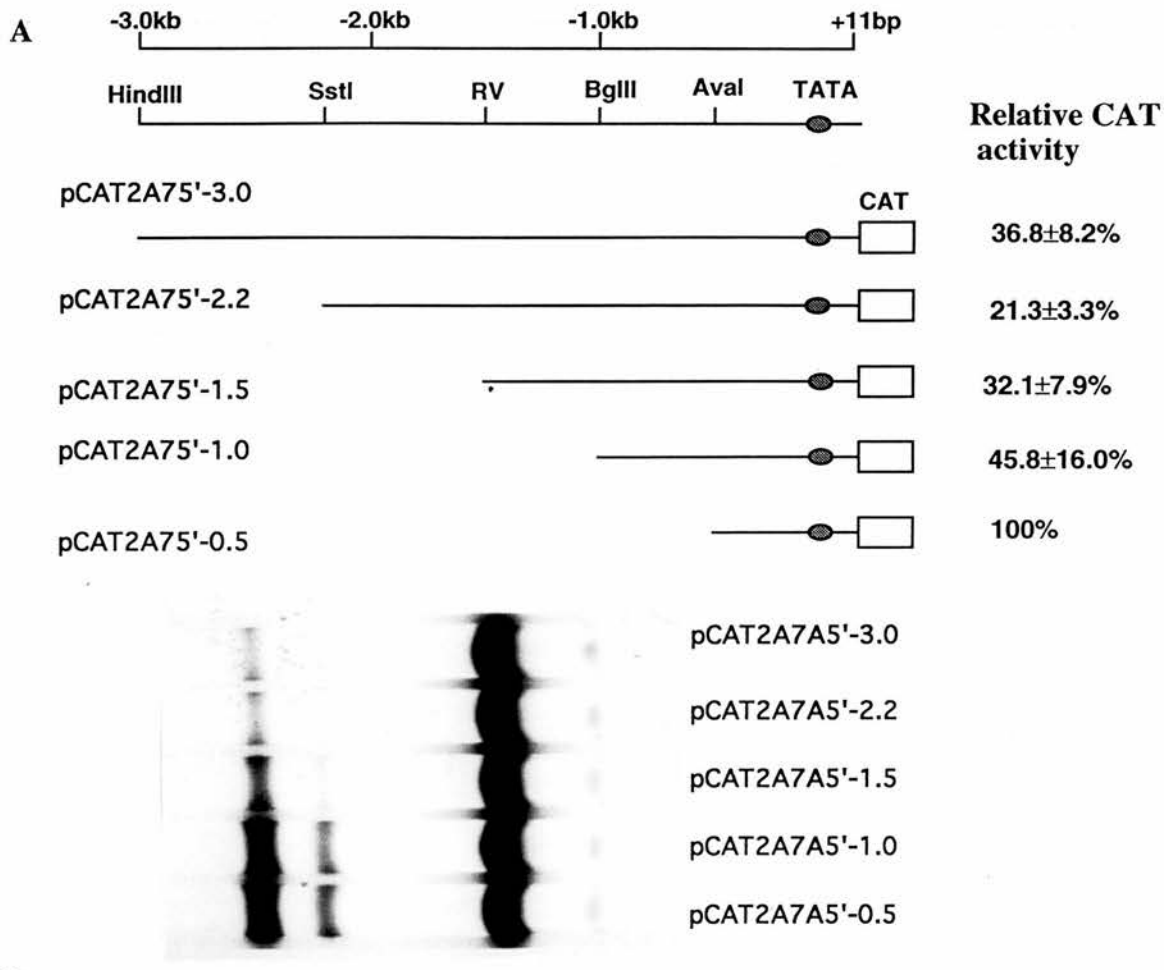
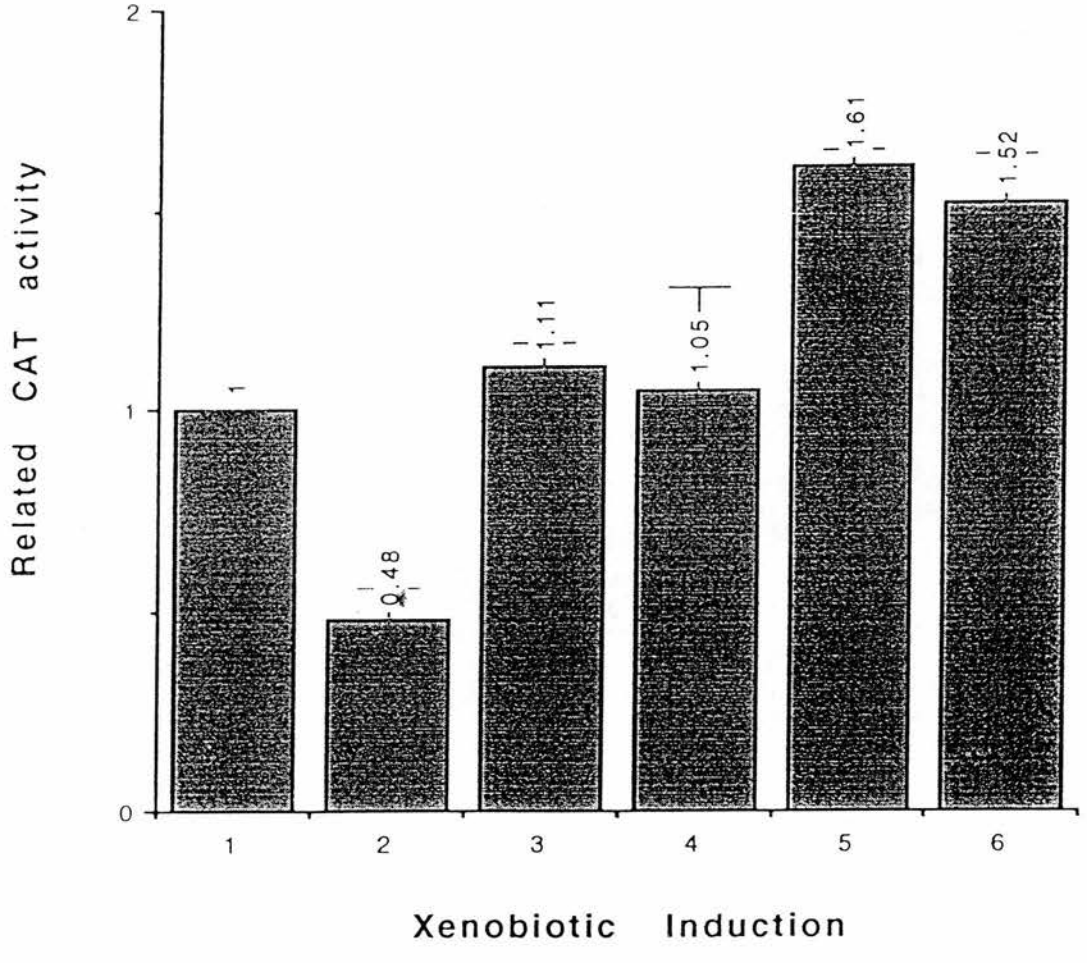

A
-3.0kb
-2.0kb
-1.0kb
+11bp
HindIII
SstI
RV
BglII
AvaI
TATA
Relative CAT activity
pCAT2A75'-3.0
CAT
36.8±8.2%
pCAT2A75'-2.2
21.3±3.3%
pCAT2A75'-1.5
32.1±7.9%
pCAT2A75'-1.0
45.8±16.0%
pCAT2A75'-0.5
100%
pCAT2A7A5'-3.0
pCAT2A7A5'-2.2
pCAT2A7A5'-1.5
pCAT2A7A5'-1.0
pCAT2A7A5'-0.5
B
-500
-400
-300
-200
-100
+11
bp
pCAT2A75'-0.5
HPF1
CAT
100%
-297bp
pCAT2A75'-D17
102.4±0.1%
-175bp
pCAT2A75'-D12
86.1±22.6%
-81bp
pCAT2A75'-D11
99.4±3.8%
-58bp
pCAT2A75'-D21
18.0±1.2%
pCAT2A7A5'-0.5
pCAT2A7A5'-D17
pCAT2A7A5'-D12
pCAT2A7A5'-D11
pCAT2A7A5'-D21

Figure 3.13 The xenobiotic induction of promoter activity. Cells transfected with pCAT2A7A5'-3.0 were grown in the presence of different xenobiotics. The CAT activity is expressed as the percentage of activity relative to the control cells. 1, control; 2, phenobarbital (2 mM); 3, TCPOBOP (20 μM); 4, ß-naphthoflavone (10 μM); 5, pyrazole (50 μM) and 6, DEX (0.1 μM). All of these chemical induction were carried out at least three times.



3.4 Discussion

Two genomic clones, CoIIA and LIIA, containing alleles of *CYP2A7*, were isolated. CoIIA contained a full length *CYP2A7A* which was approximately 8 kb in size. Although it shared several common differences with *CYP2A7* to *CYP2A6*, *CYP2A7A* contained 13 nucleotides resulting in 5 deduced amino acid differences with the *CYP2A7*. These data suggested that *CYP2A7A* appeared to be an allele of *CYP2A7*. Like *Cyp2a* genes in mice, *CYP2A7A* contained 9 exons which were divided by 8 introns at exactly the same positions as *Cyp2a4* and *Cyp2a5* .

The nucleotide and deduced sequences from exon 2 to 3 of *CYP2A7B* showed that there were eight nucleotide and six deduced amino acid differences in the region of amino acid residue 61 to 153 (nucleotide 181 to 457) between *CYP2A7B* and *CYP2A7A*, but except for two silent nucleotide changes, it was identical with *CYP2A6* in this region. On the other hand, there were four amino acid differences in the region of amino acid 158 to 164 between CYP2A7B and CYP2A6, although it was identical with that of CYP2A7A (Table 3.6). Analysis of cDNA sequences in other *CYP2* subfamilies, for example *CYP2B1* and *CYP2B2*, (Atchison and Adesnik, 1986) has revealed that regions of high nucleotide similarity are interspersed with regions of low similarity, and the mechanism underlying this phenomenon is considered to be gene conversion (Gonzalez, 1989). In the case of the human *CYP2A* subfamily, *CYP2A7A* and *CYP2A7B*, compared with *CYP2A6*, also have clusters of hypervariable regions, suggesting that gene conversion could be one of the mechanisms of *CYP2A* gene divergence during evolution. In addition, based on the PCR analysis of human genomic DNA, *CYP2A7B* appears to be another allele of *CYP2A7* (Figure 3.10). However, as only a partial sequence was available it is not clear whether *CYP2A7B* encodes a functional protein.

It has been found that expression of genes in the *CYP2* subfamily is constitutive or inducible by PB, DEX or other chemicals. In addition, tissue specific and sex-dependent regulation has also been described (Gonzalez, 1989). To date, however, little information is available about the *cis*-acting elements involved in the regulation of human *CYP2* genes and no receptor or transcription factors have been described because of lack of a suitable cell line that can express human *CYP2* genes normally. In my study it was found, by transient expression, that the deletion of the HPF-1 binding site in a fused gene construct pCAT2A7A-5'-D21 resulted in an 80% decrease of the promoter activity in HepG2 cells. A similar result was obtained in mouse HepaI cells (data not shown). Recently, it was reported that HPF-1 was highly homologous to HNF-4 and present in liver and kidney. Both factors can bind to either the HPF-1 or HNF-4 site (Chen *et al.*, 1994). In the case of the rabbit *CYP2C* gene (Venepally *et al.*, 1992), a large reduction of promoter activity in HepG2 cells is caused by deletion of the HPF-1 binding site, but this effect is rather moderate in COS-1 cells. These results show that HPF-1 is an essential element in the basal transcription of the human *CYP2A7A* in HepG2 cells and HPF-1 and HNF-4 may play a common role for the regulation of *CYP2* gene expression in humans and in other mammals. Upon treatment of the transfected HepG2 cells with various known inducers, I did not observe a significant induction by PB, DEX and TCPOBOP. This may be explained either by the absence of the specific transcription factors or receptors in the cultured HepG2 cells, or by the use of a single dosage of the chemicals, while different concentrations may have changed the induction pattern.

Computer analysis identified in the 0.5 kb flanking region of *CYP2A7A* three copies of the CACCC box, which has been found in the promoters of several genes including *CYP2H1* (Walters and Martin, 1992; Hahn *et al.*, 1991). Published results including in vitro mutagenesis and gel mobility shift assays have clearly revealed that the CACCC box plays an important role in these constitutive promoter activities and has been identified as a binding site of Sp-1, Sp-3 and Sp-4 (Yu *et al.*, 1991; Hagen *et al.*,

1992). Co-operation with the progesterone receptor (PRE) or GRE, CACCC box can enhance the transcription activity dramatically (Schüle *et al.*, 1988a).

It has been reported that SRE plays an essential role in the co-ordinate transcriptional regulation of the genes involved in the maintenance of cholesterol homeostasis and function like an enhancer (Osborne *et al.*, 1988; Smith *et al.*, 1988). Recently it has also been found that a SRE was positioned in the promoter of the human *CYP7* gene which is responsible for the metabolism of cholesterol and bile acid biosynthesis (Molowa *et al.*, 1992). Interestingly, our finding that a SRE element (CACCCCAC) overlapped with two directly repeated CACCC elements at the 5'-flanking region of *CYP2A7A* is reminiscent of cholesterol homeostasis that is related to the LDL receptor gene, in which the SRE is located between two Sp1 sites (Goldstein and Brown, 1990). Although the function of the *CYP2A7A* gene is still unclear, this finding together with the facts that murine Cyp2a proteins metabolise steroid hormones (Lindberg *et al.*, 1989a; Burkhart *et al.*, 1990) suggest that function and regulation of the *CYP2A7A* gene may be related with steroid metabolism.

CHAPTER 4: EXPRESSION AND ALTERNATIVE SPLICE OF CYP2A7

4.1 Introduction

The levels of P450s are determined by many factors, including genetic background, dietary habits, hormonal levels and exposure to foreign chemicals that act as inducers or repressors (Gonzalez, 1990). As a consequence, the enzymatic activities of hepatic P450s, particularly those in the CYP2 family, are subject to a marked inter- and intra-species variability. A significant proportion of chemical metabolising enzymes are polymorphic in man, and this results in a huge inter-individual variability of some P450 activities (Idle *et al.*, 1992). In order to define the inter-individual differences in P450 activity, it is important to delineate the underlying mechanism.

4.1.1 Polymorphisms of cytochrome P450 activities

Over the past 10 years, a great deal of information regarding the regulation of the expression and catalytic activities of the human P450s has been available and several well-defined polymorphisms in xenobiotic metabolism have been discovered. The importance of genetic polymorphisms is exemplified by studies on the genetic polymorphism at the *CYP2D6* locus. It has been demonstrated that CYP2D6 has debrisoquine 4-hydroxylase activity (Mahgoub *et al.*, 1977). The vast majority of individuals receiving debrisoquine excrete large amounts of hydroxlated metabolites in their urine (EM phenotype). About 5-10 % of the Caucasian population (Eichelbaum *et al.*, 1992a) excrete the drug virtually unchanged (PM phenotype). However, only 0.5-1.0% of Chinese and Japanese are PM phenotype (Eichelbaum and Gross, 1992b; Nakamura *et al.*, 1985), and the PM individuals are as high as 16% in Nigeria (Gonzalez and Nebert, 1990). These studies suggest that significant ethnic differences exist in chemical metabolism. In addition to debrisoquine, more than 25 other drugs are metabolised poorly by PM individuals (Meyer *et al.*, 1990), and the ability of the EM individuals to metabolise these chemicals is 10-200 times higher than that of individuals

with PM phenotype (Gough *et al.*, 1990). The PM phenotype is inherited in an autosomal recessive fashion with the EM phenotype comprising both the homozygous dominant and heterozygote genotype (Morais *et al.*, 1994). Because frequency variations of both phenotypes occur in different populations, genetic polymorphism of drug metabolism is seen as a principal determinant in inter-individual differences in toxic responses to clinical drugs and in susceptibility to chemicals.

Data reveal that the polymorphic enzymatic activity and expression of P450 can be caused by either a defect in the regulatory factors governing the transcription of the genes or direct mutations of the P450 genes. The PM phenotype has been found to be the result of defective *CYP2D6* alleles, including point mutation and deletion. In the last 5 years, more than 90% of the *CYP2D6* mutations that cause absence of CYP2D6 protein and result in the poor metaboliser phenotype have been identified by PCR/restriction fragment length polymorphism (RFLP) studies (Smith *et al.*, 1992; Daly *et al.*, 1991; Broly *et al.*, 1991; Skoda *et al.*, 1988). The most common mutant allele (>70% of PM alleles) is characterised by a point mutation at a splice-site recognition sequence that leads to a frame shift. 5% of PM individuals have a 1 bp deletion in the coding region and another 10-15% of cases are caused by deletion of the entire *CYP2D6* gene (Meyer, 1994).

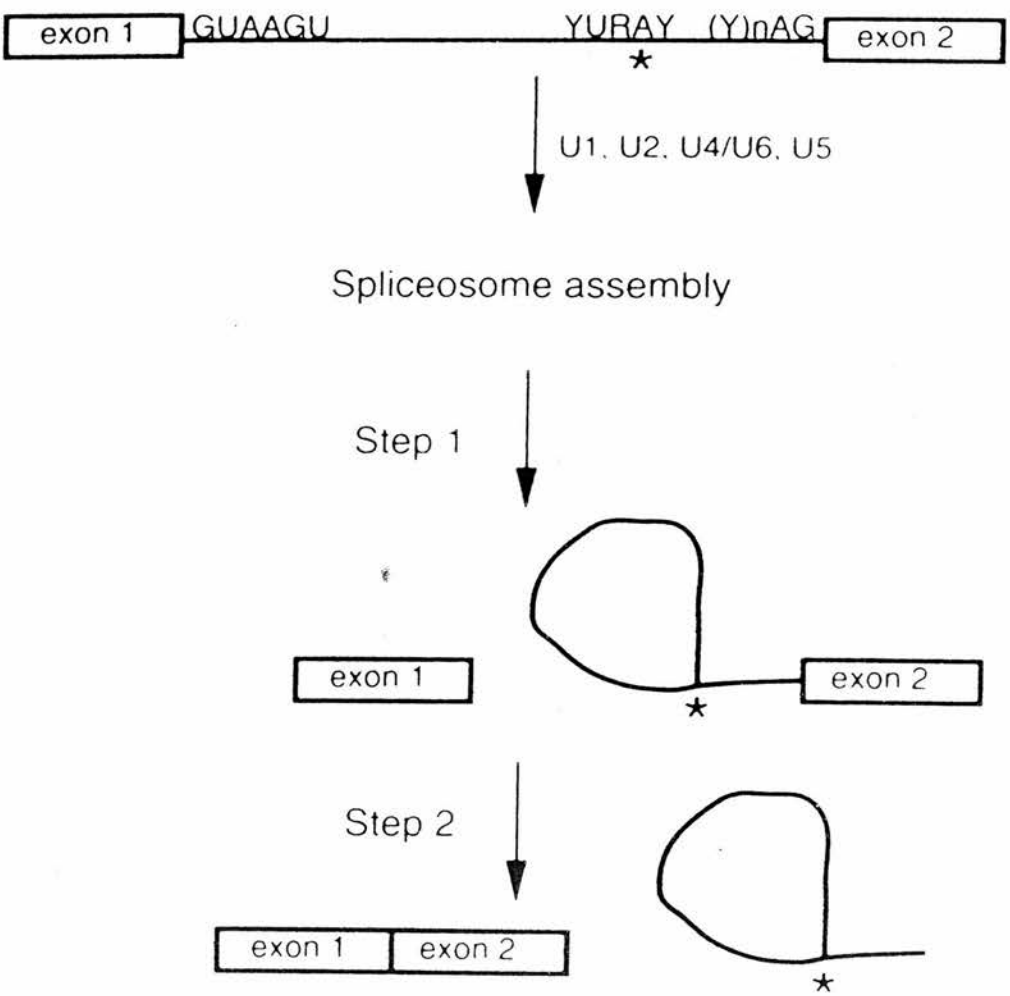
The genetic defect provides an explanation at a molecular level for the poor metaboliser phenotype. However, the mechanisms underlying the extensive metaboliser still remain unclear. Recent studies (Johansson *et al.*, 1993) have shed some light on this problem. In two families of extremely rapid metabolisers of debrisoquine, the *CYP2D6* gene was amplified and this amplification caused ultrarapid metabolism. In the three members in one family, one allele of *CYP2D6* was amplified 12-fold. Two gene copies of the same allele of *CYP2D6* existed in another family of extremely rapid metabolisers. The duplicated and amplified allele was named *CYP2D6L*. This new finding provided a molecular explanation for the rapid metabolism phenotype.

4.1.2 Alternative splice

Almost all protein-coding genes in eukaryotes are split into coding (exon) and non-coding (including intron) sequences. The intron sequences are precisely spliced out of the initial gene transcript before the mRNA is transported to the cytoplasm for translation. Each intron is removed in a two-step process. First, cleavage at the boundary between the intron and the exon upstream (the 5' splice site or splice donor) occurs, with concomitant joining of the 5' phosphate at the 5' splice site to a 2' hydroxyl of a residue within the intron. This lariat intermediate undergoes cleavage at the junction between the intron and the downstream exon (the 3' splice site or splice acceptor), with the co-ordinated ligation of the two free exons to form mature mRNA (Figure 4.1). These processes are called the general or constitutive splice pathway (Shapiro and Senapathy, 1987; Norton, 1994).

Since splicing signals are usually 5-10 nucleotides long, it is possible that a splice site is created or destroyed with an appreciable frequency by mutation or other genetic changes. This can result in the production of different polypeptides from the same gene by alternative splicing. In this case, the distinction between exons and introns is no longer absolute but depends on the mRNA reference. When the newly formed splice site is much stronger than the old one, it will have deleterious effects and is not expected ever to become fixed in the population as this process would lead to the inactivation of a previously functional protein. On the other hand, if the alternative splice site is much weaker, most mRNA will be of the original type, and only small quantities of aberrantly spliced mRNA will be made. This creates an opportunity to produce a new protein, possibly with a new useful function (Li and Graur, 1991). For example, the human insulin receptor is encoded by a single gene and composed of 22 exons. The mature insulin receptor, however, exists as two isoforms, designated A and B, which result from alternative splicing of the primary transcript. The A isoform is

Figure 4.1 Pre-mRNA splicing pathway. At the top is a sample model of pre-mRNA, with consensus elements within the intron shown: boxes, exons; line, intron; n is any residue; R is either A or G; Y is either U or C; The site of branch site formation is indicated by an asterisk. In the presence of nuclear extract, the indicated snRNPs associate with specific regions of the pre-mRNA. Formation of the higher order structure has been proposed to bring the two ends of the intron into close proximity. Splicing occurs in two steps: the characteristic intermediates and products of the reaction are shown. (Norton, P.A., 1994)



expressed only in lymphocytes, brain and spleen; the isoform B is expressed predominantly in liver, muscle, adipocytes, and kidney. It has been suggested that isoform B plays an important role in signalling in insulin-sensitive tissues (Kosaki *et al.*, 1993). Alternative splicing may be used as a means of developmental and tissue specific regulation or as a rate-limiting regulation of special genes as well (Ali *et al.*, 1992; Guo and Helfman, 1993; Zanussi *et al.*, 1992).

It is intriguing to note that alternatively spliced mRNAs have also been reported in several genes of the *CYP2* family, such as human *CYP2B*, and rat *CYP2C* (Miles *et al.*, 1988; Okino *et al.*, 1987; Kimura-H. *et al.*, 1989; Zaphiropoulos, 1993). The alternative splicing has led to inactive protein products. For example, two types of mRNA are transcribed from one rat *CYP2C6* gene by alternative splicing in exon 8 and both of the two mRNAs are expressed in rat liver in an age-dependent manner. The 142 bp exon 8 in the wild type transcript of *CYP2C6* is replaced by a 159 bp fragment, resulting in a disruption of the open reading frame. Although the aberrantly spliced mRNA can be translated into a protein, its haem-binding capacity is lost and the protein does not function as a P450 monooxygenase (Kimura-H. *et al.*, 1989). Recently, it has been reported (Morais *et al.*, 1994) that a principle defect in the poor metaboliser phenotype of S-mephenytoin hydroxylation was a single base pair mutation in exon 5 of *CYP2C19*, which creates an alternative splice site. The mutation alters the reading frame of the mRNA and produces a premature stop codon, which results in a non-functional protein in man.

Another example is the expression of the human *CYP2B6* gene. The investigation on this gene, whose expression has been found to be co-regulated with the *CYP2A* gene, indicates that three types of transcripts from the gene are formed by aberrant splicing in exon 8 in all the 15 tested human liver samples, but their ratios vary among the samples (Miles *et al.*, 1988). The alternatively spliced mRNAs from the *CYP2B6* gene generate a protein product without activity. Taken together, these results suggest that alternative

splicing of mRNA should be one of the principal regulating mechanisms in the *CYP2* gene family, and appears to be one of the molecular explanations for inter-individual variation in the activity and expression of P450.

4.1.3 Research purpose

In order to determine the role of P450s in individual adverse drug reactions and chemical toxicity, the basis for polymorphisms in gene expression needs to be established. In this regard the genes of the human *CYP2A* subfamily have not been extensively investigated. Both genes *CYP2A6* and *CYP2A7* are expressed in human liver, but only CYP2A6 has been found to be responsible for the metabolism of coumarin (Yamano *et al.*, 1990; Miles *et al.*, 1990) as well as the carcinogen aflatoxin B_1 (Crespi *et al.*, 1991). The functions of CYP2A7 are currently unknown.

The expression level of CYP2A6 is highly variable within the population. Coumarin hydroxylase (COH) activity varies greatly as well (up to 100-fold, Yun et al., 1991). This appears to be consistent with the difference in hepatic CYP2A expression in previously published results (Pelkonen *et al.*, 1993; Yamano *et al.*, 1990). Northern blot analysis with *CYP2A6* cDNA as a probe identifies two mRNAs of 2.3 and 2.8 kb in human liver, and immunoblotting of liver microsomes also reveals two or three bands, 49 kDa, 51 kDa and 55 kDa (Yamano *et al.*, 1990; Miles *et al.*, 1990; Forrester *et al.*, 1992). These results suggest that more than two *CYP2A* genes may be expressed in human.

In order to study the factors involved in the inter-individual variability at the expression levels of human *CYP2A* genes, the expression levels of *CYP2A* genes in human liver was studied using RT-PCR. Three mRNA species encoding CYP2A6, CYP2A7 and an alternatively spliced form of CYP2A7 have been subcloned and isolated in this work. The identification of the latter mRNA exemplifies previous reports indicating that

alternative splicing is an important factor in determining cytochrome P450 levels in man.

4.2 Results

4.2.1 Determination of CYP2A mRNA levels in human liver

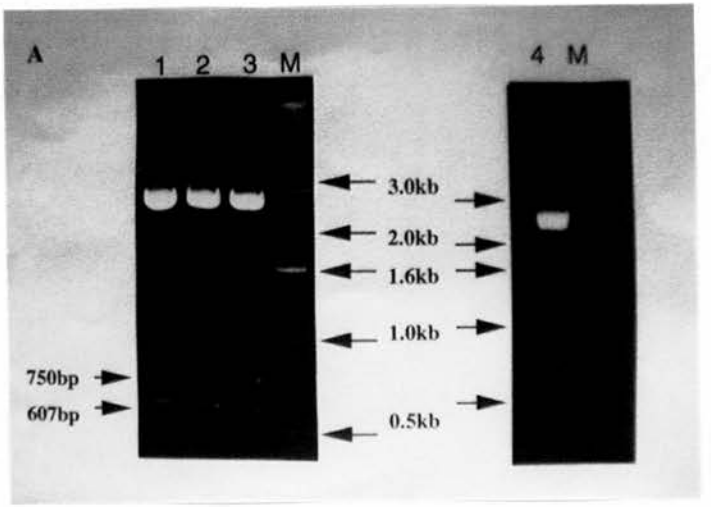
Analysis of the human *CYP2A6* and *CYP2A7* cDNA sequences (Miles *et al.*, 1990) showed the presence of a *Pst*I restriction site at bp 143 (+1 indicates the start of the open reading frame) of *CYP2A6*, which is absent in *CYP2A7*. This restriction enzyme *Pst*I digestion combined with a protocol of RT-PCR was employed to establish the relative mRNA levels encoded by these two genes. To amplify a fragment spanning exons 1 to 5 of the *CYP2A6* and *CYP2A7* genes, an upstream oligonucleotide A, 5'-CATGCTGGCCTCAGGGCTGCTT-3' and a downstream oligonucleotide C, 5'-GAAGTCCTCCAGCCCTTGCAGC-3', were used. The human livers employed in this determination were obtained from kidney transplant donors. Livers were stored at -70^{0}C within 1 h of removal. Information about the patient case histories have been described previously (Miles *et al.*, 1990; Forrester *et al.*, 1992). Total cellular RNA was prepared by the guanidinium isothiocyanate method or a single-step method described in chapter 2. RNA concentration and purity were estimated spectrophotometrically. Before reverse transcription, RNA was tested for integrity by ethidium bromide staining following separation on a denaturing formaldehyde gel.

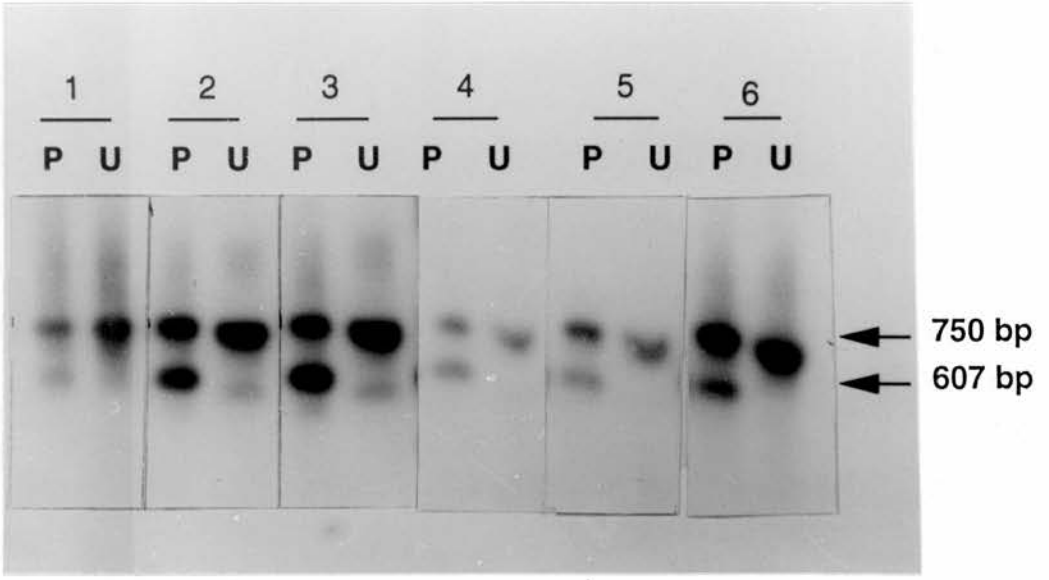
First, hepatic mRNA was reverse transcribed using a oligo $(dT)_{6-8}$ and reverse transcriptase (MuLV from BRL Ltd.), and then CYP2A6 and CYP2A7 were amplified by PCR using *Taq* polymerase (Promega Ltd.). The expected 750 bp fragments of *CYP2A6* and *CYP2A7* were observed (Figure 4.2, the tracks marked by U). Following digestion with *Pst*I (20 units and four hours incubation at 37^{0}C), *CYP2A6* cDNA was as expected cut into two fragments of 607 bp and 143 bp, whereas *CYP2A7* cDNA was unaffected by this procedure. During subsequent electrophoresis

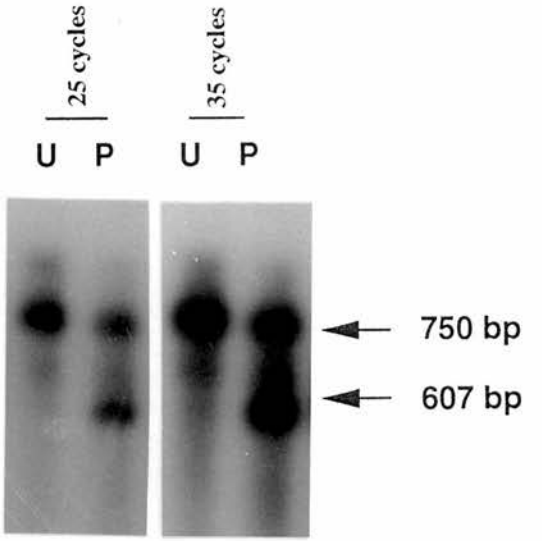
the 143 bp fragment migrated out of the gel and was not seen. In a separate experiment, 0.4 μg pBluescript plasmid DNA was added into the reaction mixture as an internal control (Figure 4.2A) to verify complete digestion. In order to ensure that the analysis was reproducible, the RT-PCR followed by *Pst*I digestion for samples L8, L6 and L4 were carried out twice. The results showed that the ratio of CYP2A7 to CYP2A6 in these three RNA samples were reproducible (results not shown). Both *CYP2A6* and *CYP2A7* were expressed in all these liver samples. The ratio of the expression level of *CYP2A7* to that of *CYP2A6* was subject to some variation ranging from slightly less than one in sample 3 to about three to four in (Figure 4.2B). The relative expression of

Figure 4.2. The relative amounts of CYP2A6 and CYP2A7 mRNA in six human liver samples. (**A**) Agarose gel electrophoresis of *Pst*I digested RT-PCR products. 0.4 μg plasmid pBluescript DNA (2.94 kb) was added into the reaction mixture to verify complete digestion. M, 1 kb ladder marker; lanes 1 to 3, digested pBluescript DNA (2.94 kb) with RT-PCR products from RNA samples L8, L9 and L11, respectively; 4, uncut pBluescript DNA. (**B**) Southern blot analysis of *Pst*I digested RT-PCR products. Number 1 to 6 represent the RNA samples L12, L8, L6, L16, L14 and L4, respectively. The fragment of 607 bp represents the digested *CYP2A6* and that of fragment of 750 bp is *CYP2A7*. (**C**)The ratio of relative expression of the *CYP2A7* to *CYP2A6* over a range of PCR cycles with RNA sample L6; lane U, uncut PCR product; lanes P, *Pst*I digested product. The designation of the samples is the same as that described in literature (Miles *et al.*, 1990; Forrester *et al.*, 1992).

Figure 4.2


A
1 2 3 M
4 M
3.0kb
2.0kb
1.6kb
1.0kb
0.5kb
750bp
607bp


B
1 2 3 4 5 6
P U P U P U P U P U P U
750 bp
607 bp


C
25 cycles
35 cycles
U P U P
750 bp
607 bp

the *CYP2A7* to *CYP2A6* in one RNA sample (L6) was determined over a range of PCR cycles. The amplified DNA was cut with *Pst*I and the results showed that the expression level of *CYP2A7* to that of *CYP2A6* was constant with the number of PCR cycles (Fig. 4.2C).

4.2.2 The identification of an alternatively spliced CYP2A7 mRNA

As part of the analysis of CYP2A6 and CYP2A7 mRNA's, RT-PCR was carried out to amplify the full length cDNA of *CYP2A* genes using a human total liver RNA sample L8 and oligonucleotide primers. The primer A was complementary to the 5' end, bp 1-22, (See section 4.2.1). The primer B was complementary to the 3' end, bp 1583-1564 (5'-GCCTTAAGGCTTCCCCCCATTCTTATACC-3', containing an additional *Eco*RI site). In addition to the expected fragment of 1.6 kb, a low abundance fragment of about 1.45 kb was observed. The PCR products were subcloned into the *Sma*I/*Eco*RI sites of vector pUC19. *Pst*I digestion was used to screen the colonies as this site is present at bp 143 (+1 indicates the start of the open reading frame) of *CYP2A6* , but absent in *CYP2A7*. Three colonies with different *Pst*I digestion patterns were isolated and sequenced. The results indicated that the colonies with a 1.6 kb insert contained a *CYP2A6* or a *CYP2A7* cDNA. The colony with a 1.45 kb insert appeared to contain an alternatively spliced version of *CYP2A7* (*CYP2A7AS*).

Comparing the genomic sequence of *CYP2A7* gene with that of *CYP2A7AS*, it was confirmed that the *CYP2A7AS* was an alternatively spliced version of *CYP2A7* in which the 163 bp of exon 2 was replaced by a 10 bp segment of intron 1 (Figure 4.3A). The alternatively spliced exon 1 was designated exon 1A. Translation of CYP2A7AS mRNA would result in an in-frame deletion of 51 amino acids to generate a protein product of M_r 44 kDa. The 10 bp segment of intron 1 added three amino acid residues at amino acid 60 and was then linked to amino acid 114 in exon 3

Figure 4.3A Structure of the normal and alternatively spliced CYP2A7 mRNAs. Exon 1A contains exon 1 plus the first 10 bp of intron 1.

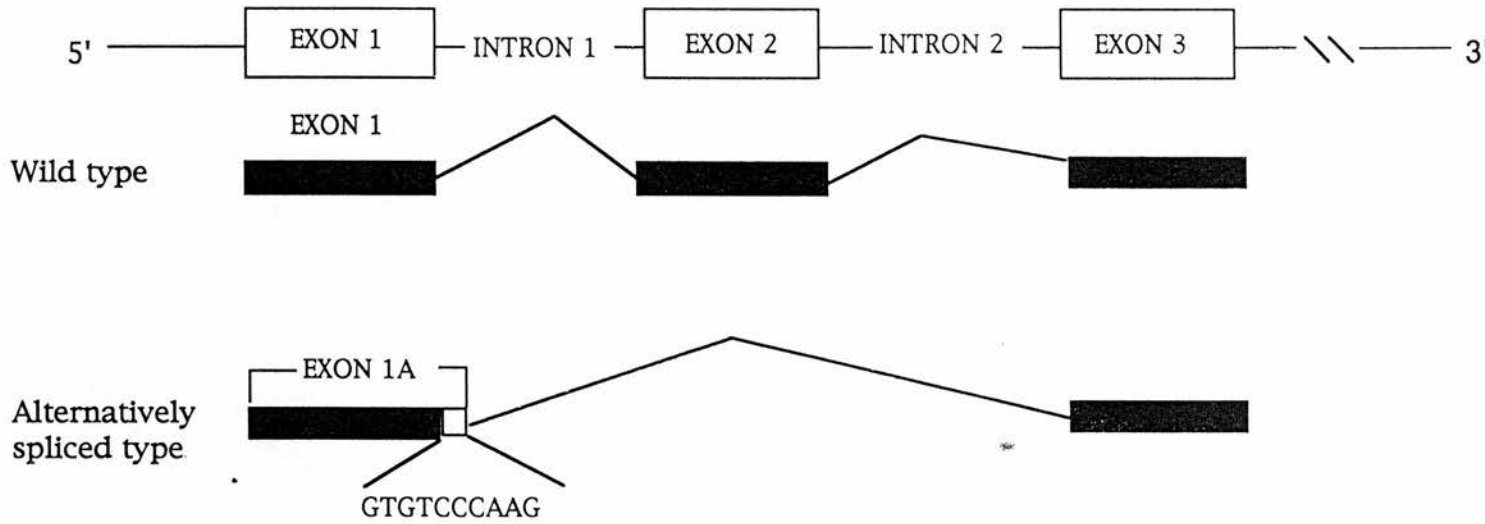


Figure 4.3B Deduced amino acid sequences of *CYP2A7* and alternatively spliced *CYP2A7* (*CYP2A7AS*). In CYP2A7AS the 54 amino acids of exon 2 were replaced by three amino acid residues Val-Ser-Gln.

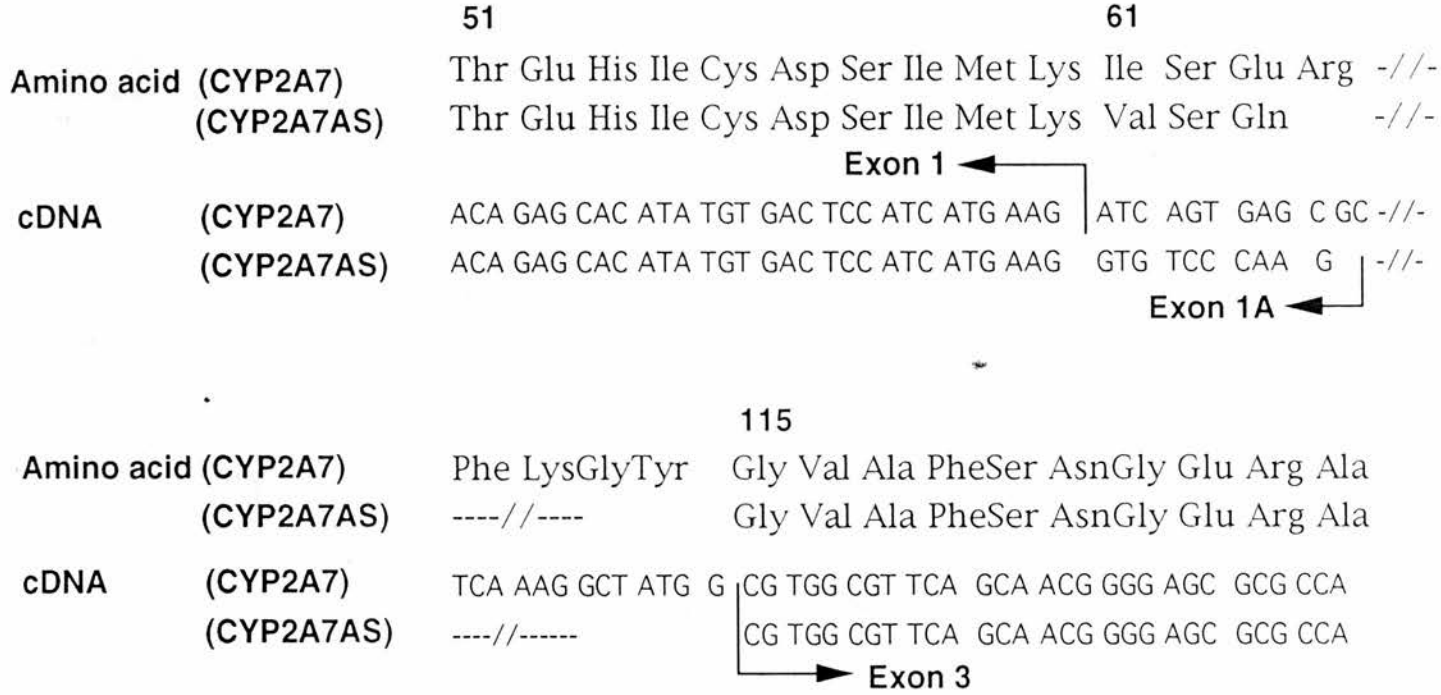
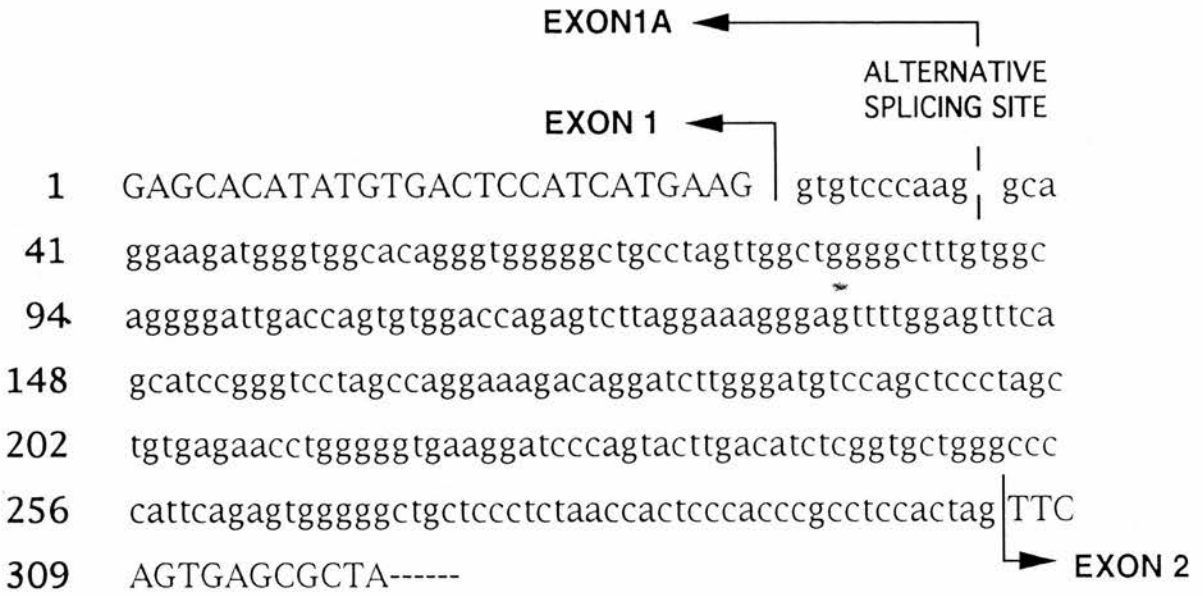


Figure 4.4 Sequence covering the alternatively spliced region of *CYP2A7*. The alternative site is indicated by a vertical broken line. The exon sequences are designated by upper case; intron sequences by lower case letters.



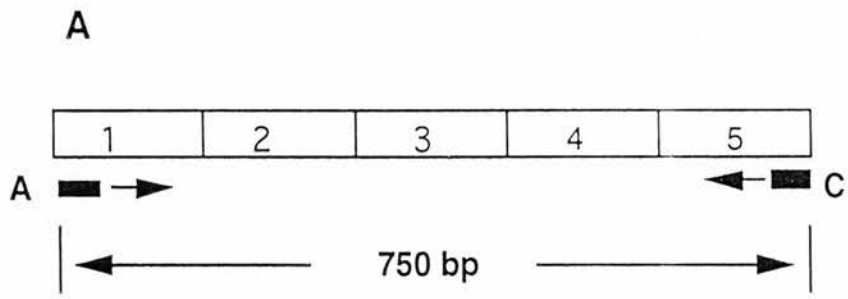
(Figure 4.3B). All the intron/exon junctions conformed to the GT/AG consensus splice recognition site. However, intron 1 contained an additional 5'-splice site, G/gcagg, which resulted in CYP2A7AS (Figure 4.4). This alternative 5'-splice site did not conform to the GT/AG consensus sequence, however, it was similar to a "non-conforming" 5' splice site sequence found in a few genes (Shapiro and Senapathy, 1987; Miles *et al.*, 1989b).

4.2.3 The expression of the alternatively spliced CYP2A7 in human liver and a skin fibroblast cell line

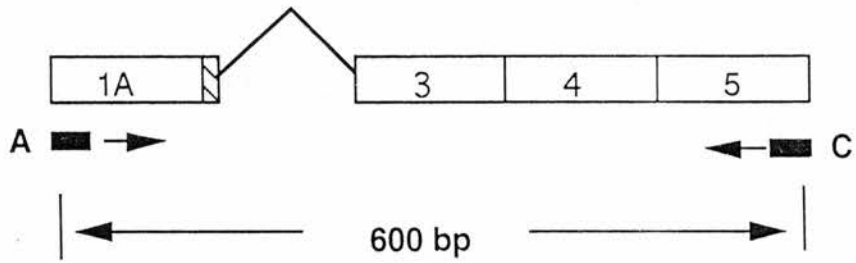
To investigate the extent of alternative splicing of CYP2A7, five human liver RNA samples and total RNA from cultured human skin fibroblast cells were assayed by RT-PCR. The expected wild type PCR product using oligonucleotide A and C as primers is a fragment of 750 bp. The product from the alternatively spliced mRNA is 600 bp (Figure 4.5A). An ethidium bromide stained band of 750 bp appeared in four of five liver RNA samples, whereas a weak band of 750 bp and a much stronger 600 bp band existed in the PCR product of the skin fibroblast cell RNA (Figure 4.5B). After transferring to a Hybond-N nylon membrane and hybridising with a 0.78 kb probe (spanning exon 1 to exon 5 of *CYP2A7*), four of five samples, including sample L5

Figure 4.5 Analysis of CYP2A7 species. (**A**) Analysis procedure: RT-PCR was carried out with primers A and C. (**B**) The products were analysed by ethidium bromide staining. Lane 1, human fibroblast total RNA. Lanes 2-6, Human liver RNA samples L14, L15, L8, L6 and L4, respectively. (**C**), Southern hybridisation of the RT-PCR products with a 0.78 kb cDNA probe containing exon 1 to 5 of *CYP2A7*. Lane 3 underwent a longer exposure.

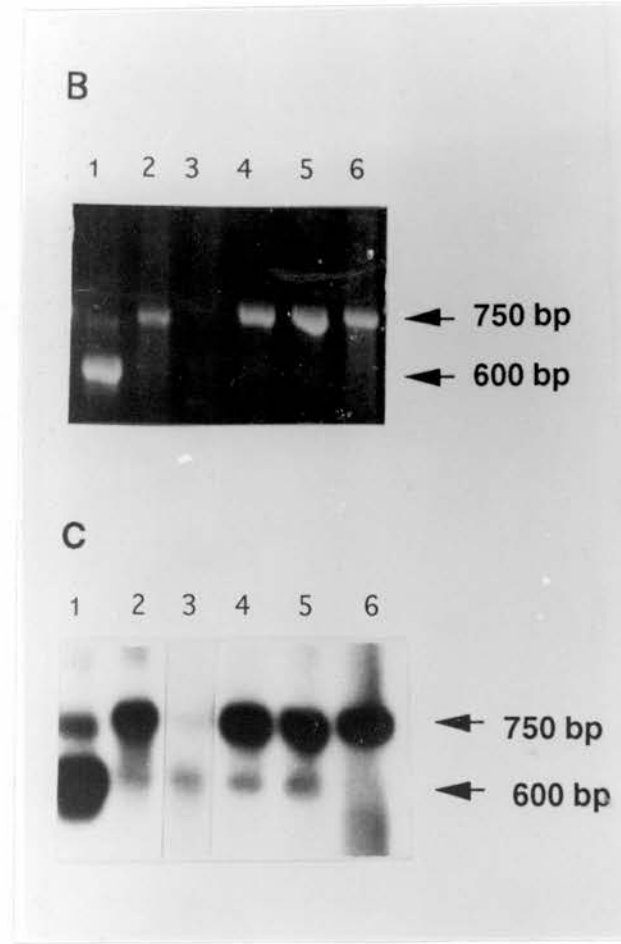
Figure 4.5



PCR product of wild type mRNA



PCR product of alternatively spliced mRNA



(after longer exposure), contained alternatively spliced CYP2A7 mRNA. Variability in the relative level of this mRNA vs. normal CYP2A7 mRNA was found between samples (Figure 4.5C). Indeed, in sample L5 the level of the alternatively spliced mRNA was 3 to 4-fold higher than the correctly spliced transcript. However in the other samples, the normal transcript was the predominant mRNA species. Interestingly, the major CYP2A7 mRNA species in the skin fibroblast cell line appears to be the alternatively spliced mRNA, with only a very low amount of the normal transcript (Figure 4.5B and 4.5C, track 1).

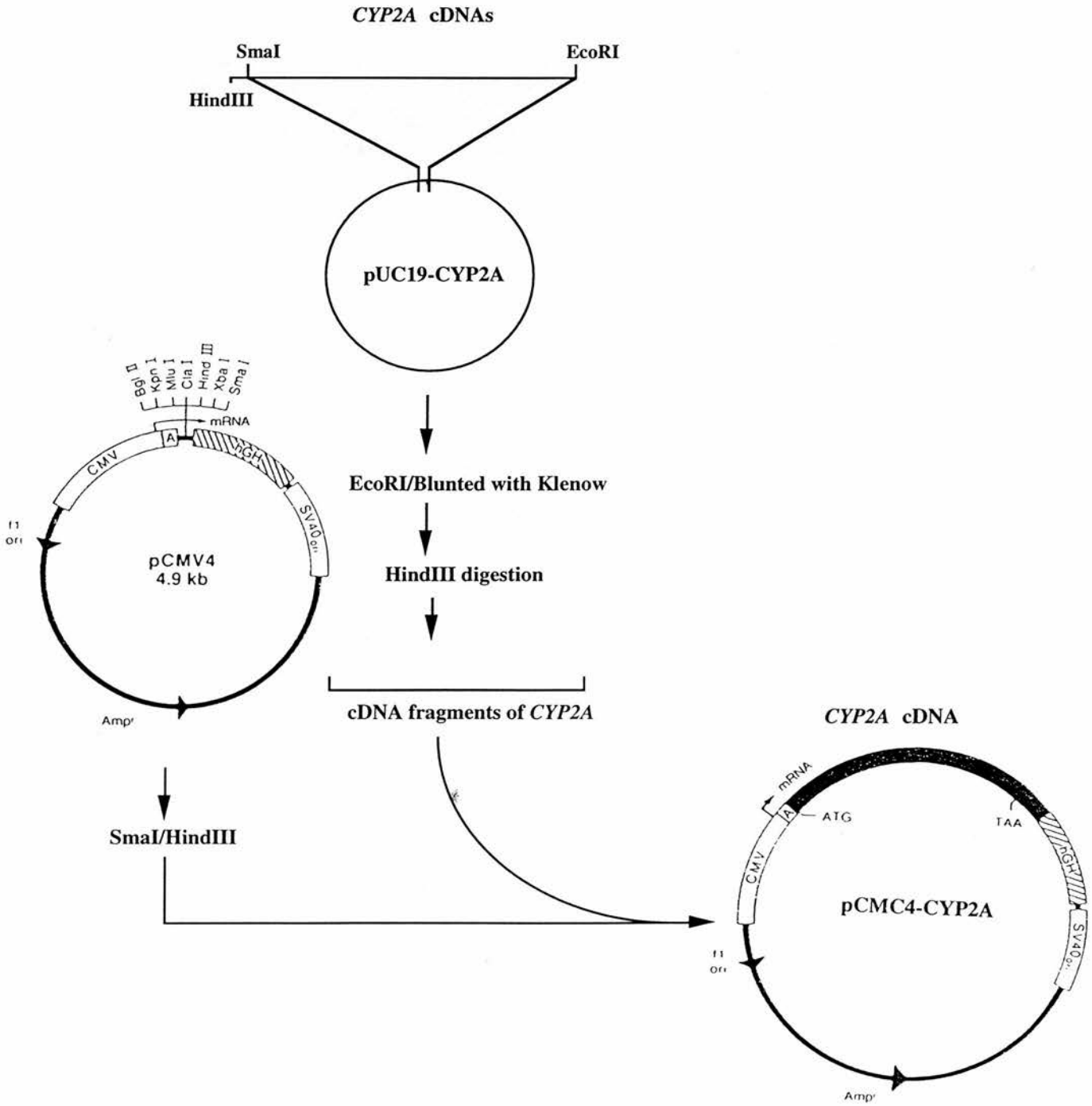
4.2.4 Functional analysis of CYP2A transcripts

To investigate whether these mRNAs could be translated, the *CYP2A6*, *CYP2A7* and alternatively spliced *CYP2A7* cDNAs were subcloned separately into a vector pCMV4. The pCMV4 is a mammalian expression vector which contains the promoter-enhancer sequence of the major immediate early gene of human cytomegalovirus, and the polyadenylation region of the human growth hormone gene. The SV40 origin of replication in $pCMV_4$ can increase transfectional efficiency in COS cells which had been previously transformed with large T antigen of SV40 (Andersson *et al.*, 1989)

The subcloning strategy is as follows: the full length *CYP2A6* cDNA was described previously by Miles *et al.* (1990). The subcloned *CYP2A6*, *CYP2A7* and *CYP2A7AS* cDNAs in vector pUC19 were digested with *EcoR*I, blunted using Klenow enzyme and then cut with *Hind*III. The resulting fragments were then ligated into the *Sma*I-*Hind*III sites of $pCMV_4$, separately, to form the expression plasmids $pCMV_4$-CYP2A6, $pCMV_4$-CYP2A7 and $pCMV_4$-CYP2A7AS (Figure 4.6). These constructs were used to transiently transfect COS cells. The transfected COS cells were lysed by sonication using an MSE Soniprep (two 5 second bursts at an amplitude of 12 mm with sample kept on ice). The lysed samples were centrifuged at 13000 rpm for 10 min to prepare crude supernatant and pellet fractions for Western blot analysis. Proteins were

Figure 4.6 Structure of human CYP2A expression vector. Plasmid $pCMV_4$ represents the starting expression vector and contains the immediate early promoter region of human cytomegalovirus (CMV, stippled block), a DNA copy of a segment of the alfa mosaic virus 4 RNA that contains a translational enhancer (A), a polylinker containing unique sites for the indicated restriction enzymes, transcription termination and polyadenylation signals from the human growth hormone gene (hGH, hatched block), and the SV40 origin of DNA replication and early region enhancer sequences ($SV40_{ori}$, white block). The plasmid also contains an *E. coli* gene encoding ampicillin resistance (Amp^r) and a bacteriophage f1 origin of DNA replication (f1). Plasmid pUC19-CYP2A represents the plasmids containing inserts of the *CYP2A6* or *CYP2A7* or *CYP2A7AS* . Plasmid $pCMV_4$-CYP2A contains an inserted fragment of the human CYP2A6 or CYP2A7 or CYP2A7AS cloned into the *Sma*I and *Hind*III sites of the polylinker region of $pCMV_4$. The approximate position of the initiator methionine (ATG) and translation termination (TAA) codons are indicated below the CYP2A cDNA.

Figure 4.6


CYP2A cDNAs
SmaI
EcoRI
HindIII
pUC19-CYP2A
EcoRI/Blunted with Klenow
HindIII digestion
cDNA fragments of *CYP2A*
SmaI/HindIII
pCMV4
4.9 kb
CMV
mRNA
Amp
CYP2A cDNA
ATG
TAA
pCMC4-CYP2A

separated by SDS/PAGE (Laemmli, 1970), transferred to nitro-cellulose membrane and probed with anti-rat CYP2A antiserum using the method of Lewis et al. (1988). Western blot analysis of the membrane fractions showed that all three cDNAs gave protein products (Figure 4.7A, tracks 2 to 4). The molecular weight of the CYP2A6 and CYP2A7 proteins was identical, M_r 49 kDa, with an identical mobility to the major immunostained band identified in human liver microsomes with anti-rat CYP2A1 antibody. The alternatively spliced CYP2A7 gave two protein bands, M_r 44 kDa and 42 kDa (Figure 4.7A, track 4). Since abnormal mRNA often leads to an unstable protein product, the 42 kDa protein could be a degradation product. A protein band which co-migrated with the 44 kDa protein was also found in the crude membrane fraction of the human skin fibroblast cells, but not in the human liver microsomal sample. No detectable 49 kDa protein was observed in the fibroblast cell line (Figure 4.7A, track 5).

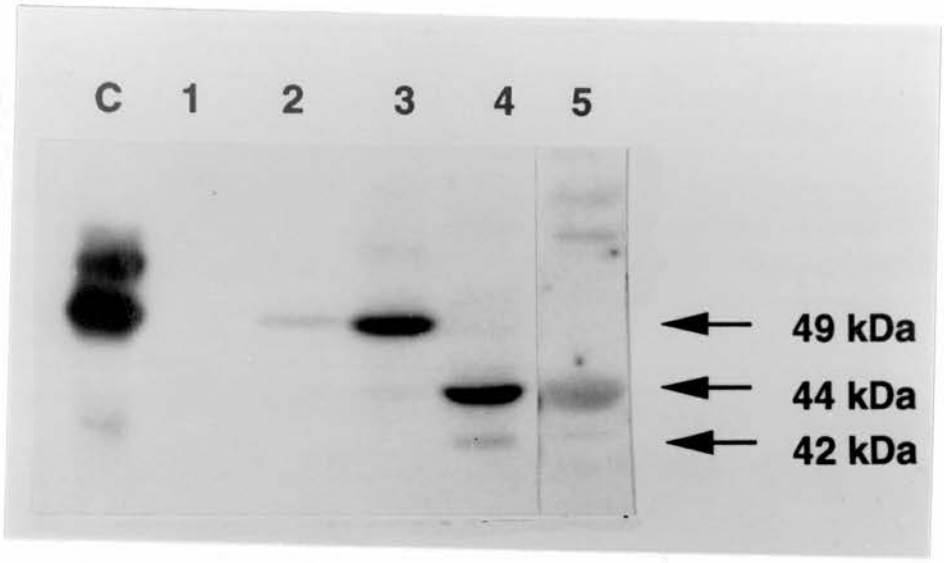
In order to determine whether the expressed proteins are catalytically active, coumarin hydroxylase activity was measured by HPLC in the transfected cells as follows. The transfected COS cells were washed with PBS and then re-fed with 5 ml fresh, serum free medium. 25 µl of 10 mM coumarin in DMSO and 10 µl of (3-^{14}C) labelled coumarin (13.25 µCi in 0.5 ml DMSO) were added into the flask and then the cells were cultured for 6 h at 37^{0}C. After incubation the medium was collected and an equal volume of ice cold methanol was added. The various polar products including 3-,4- and 7-hydroxycoumarins were determined in the medium by HPLC analysis (Iersel *et al.*, in press). The 7-hydroxylcoumarin was considerable higher in the cells expressing CYP2A6 than the controls. However, the 7-hydroxylcoumarin and the total polar products in the cells transfected with the different forms of CYP2A7 did not change (Figure 4.7B).

Figure 4.7. A. Expression of CYP2A cDNAs in COS-7 cells. cDNAs encoding CYP2A6, CYP2A7 or CYP2A7AS were subcloned into pCMV4 and transfected into COS-7 cells as described in the Materials and Methods. Western blotting analysis of proteins from the crude pellet fractions of transfected COS cells (100 μg) and of cultured human skin fibroblast cells (100 μg) was done using an antibody to rat CYP2A1. Human liver microsomal protein (10 μg, lane C) was used as a positive control. Track 1, untransfected COS-7 cells; track 2, pCMV4-CYP2A6; track 3, pCMV4-CYP2A7; track 4, pCMV4-CYP2A7AS and track 5, human skin fibroblast cells (in the same Western blotting analysis).

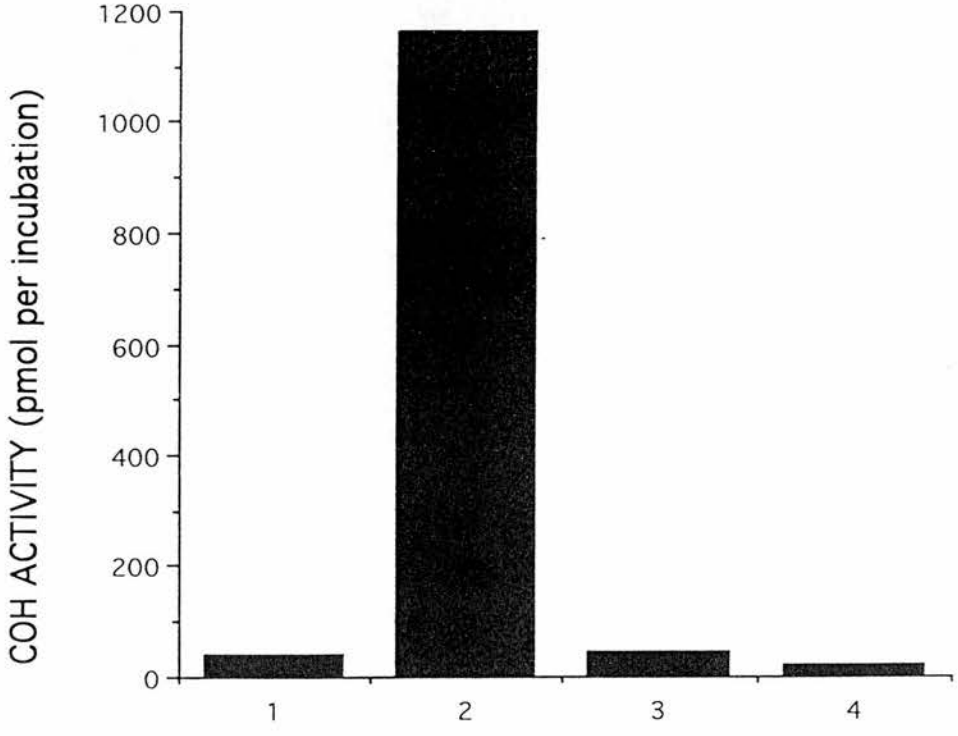
B. Coumarin-7-hydroxylase activity. A proportion of the cells from the same samples analysed by Western blot analysis were assayed for coumarin hydroxylase activity as described in the Materials and Methods by HPLC. Coh activitie is expressed as pmol/incubated (5 x 10^6 cells)/6 h. Tracks are: 1, untransfected cells; 2, pCMV4-2A6; 3, pCMV4-2A7; 4, pCMV4-2A7AS. The HPLC was done by Dr Lake from BIBRA Toxicology International and the assay can determine various polar products including 3- and 4-hydroxylcoumarins.

Figure 4.7

A


C
1
2
3
4
5
49 kDa
44 kDa
42 kDa

B


COH ACTIVITY (pmol per incubation)
1200
1000
800
600
400
200
0
1
2
3
4

4.3 Discussion

CYP2A7 mRNA can be alternatively spliced to give a transcript missing exon 2 but containing an additional three amino acids derived from intron 1. This alternatively spliced mRNA can be translated to give a protein product of the predicted molecular weight in cDNA directed expression and in a human fibroblast cell line. The relative hepatic level of the CYP2A7AS versus CYP2A7 mRNA varied considerably between individuals and in one case CYP2A7AS was the more abundant of the two mRNA species. In the human fibroblast cell line both CYP2A7 transcripts were detected, but the major product was CYP2A7AS.

It is intriguing that aberrantly spliced mRNAs have also been reported for several other genes in the *CYP2* family. For example, two mRNAs are derived from the rat *CYP2C6* gene by alternative splicing in exon 8. This like many other examples leads to a disruption of the open reading frame. Although the alternatively spliced mRNA can be translated into a truncated protein, its haem-binding capacity is lost and therefore the protein cannot function as a P450 monooxygenase (Kimura-H. *et al.*, 1989). The transcript of human *CYP2B6*, whose expression may be co-regulated with *CYP2A* (Miles *et al.*, 1989a), is also alternatively spliced, and at least four mRNA species are derived from this gene (Miles *et al.*, 1989b). The related levels of different mRNAs are subject to considerable inter-individual variability. Similar to the findings here, a variant of CYP2B6 is generated using a cryptic "non-conforming" 5'-splice site, G/gcaag. Alternative splicing has also been described for human *CYP2D* genes (Gonzalez *et al.*, 1988; Gough *et al.*, 1990). Taken together, these results suggest that alternative splicing is an important determinant in the expression of many P450 genes and this effect will contribute to the inter-individual variation in the enzyme levels. In addition, alternative splicing using cryptic non-GT conforming 5'-splice site is considered as a rate-limiting regulation for some genes (Shapiro and Senapathy, 1987).

Cytochrome P450 is synthesised on membrane-bound polysomes (Bar-Nun *et al.*, 1980) and inserted into the endoplasmic reticulum membrane using the first 29 N-terminal amino acids as an anchor (Sakaguchi *et al.*, 1987). The C-terminus of cytochrome P450 is important for haem -binding. The protein product of alternatively spliced CYP2A7 still contains the conserved P450 haem binding region and could conceivably still function as a monooxygenase enzyme. However, CYP2A7AS does not contain exon 2 which might form a potential transmembrane domain (Nelson and Strobel, 1988) and contains specific amino acids responsible for substrate recognition (Gotoh, 1992). The results of cDNA directed expression in COS cells showed that the alternatively spliced *CYP2A7* produces a truncated protein of *M*r 44 kDa and this protein was associated with the crude membrane fraction (Figure 4.7A). No detectable CYP2A7AS protein was found in the crude supernatant fraction (results not shown). This result agrees with recent membrane topology models of P450 suggesting that only exon 1 of P450s codes for the membrane anchor (Black, 1992). Since the expression level of CYP2A7AS in COS-7 cells was too low to establish whether the truncated protein still binds haem or not, studies to establish whether this is the case using other expression systems are in progress.

Studies on the CYP2A subfamily in mouse and rat have revealed the evolution and regulation of its enzyme activity. Though there are species differences in gene expression and in the number of genes, the enzymes in the CYP2A subfamily exhibit a conserved coumarin 7-hydroxylation activity. The CYP2A6 enzyme is reported to be the major enzyme catalysing coumarin 7-hydroxylation in human liver (Yamano *et al.*, 1990; Miles *et al.*, 1990), and the level of enzyme activity varies with different livers (Maurice *et al.*, 1991). This conclusion was based on the observed correlation between the level of CYP2A6 protein and the enzyme activity in human livers. Although in some tested human liver samples the intensity of the stained band in the immunoblotting assay appears to correlate with the level of coumarin 7-hydroxylase activity (Pearce *et al.*, 1993; Yamano *et al.*, 1990), some other tested samples in their reports do not show

such correlation. In my experiments cDNA directed expression of CYP2A6 and CYP2A7 in COS-7 cells indicated that CYP2A7 had exactly the same mobility as CYP2A6 on SDS-PAGE and both protein products were 49 kDa (Figure 4.7A) Since only CYP2A6 had coumarin 7-hydroxylase activity and the function of CYP2A7 has not yet been demonstrated, my finding revealed that the amount of 49 kDa protein did not seem to correlate, at least in some livers, with the level of coumarin 7-hydroxylase activity because the protein band of 49 kDa in immunoblotting assay should be a mixture of CYP2A6 and CYP2A7 products.

Immunohistochemical techniques have been used to identify the presence of P450 in skin. Enzymes of the CYP2 and CYP3 families were found in cultured rat skin cells and in normal and transformed human epidermal keratinocyte cell lines as well (Hotchkiss, 1992). The activities of P450s in skin cells are much lower than those in liver, ranging from 0.1 to 27 % (Hotchkiss, 1992). In my experiment, a high level of the alternatively spliced CYP2A7 mRNA, compared with the normal mRNA, existed in cultured human skin fibroblast cells. Western blot analysis using anti-rat CYP2A antiserum showed that there was a weakly-staining 44 kDa protein band representing the product of alternatively spliced CYP2A7 mRNA in the membrane fraction of the skin cells, but there was no detectable 49 kDa protein which was the product of CYP2A7 mRNA. This evidence clearly demonstrated that CYP2A7 gene was not only expressed in liver, but also in cultured skin cells.

The mechanism of alternative splicing of CYP2A7 mRNA is not known. The basis for the variability in alternative splicing of CYP2A7 mRNA could be determined by either genetic and/or environmental factors. The results presented in this chapter suggest that the alternate splicing process may be an important determinant in the expression of *CYP2A7* genes and play a rate-limiting regulation for this gene. The latter may be related to the inter-individual variation of CYP2A enzyme levels. Since a high level of alternatively spliced CYP2A7 was observed in cultured skin cells and in one of the

tested livers in my experiment, environmental factors or *in vitro* culture conditions may play a function in the alternative splicing of CYP2A7 mRNA as well.

CHAPTER 5: SUMMARY AND FUTURE WORK

In this thesis, the structure and regulation of the human *CYP2A7A* gene have been described. The results of genomic DNA cloning and PCR/RFLP analysis indicated that there were two alleles of *CYP2A7* (*CYP2A7A* and *CYP2A7B*) in humans. In addition, based on the results of expression levels of *CYP2A* genes in human livers and alternative splicing of CYP2A7 mRNA, the possible mechanisms of the inter-individual variability of coumarin 7-hydroxylase (Coh) activity have been discussed in detail.

5.1 The Structures of *CYP2A7* Alleles

Two genomic clones, CoIIA and LIIA, containing alleles of human *CYP2A7* were isolated. The clone CoIIA, isolated from a human cosmid library, contained a full length version of the *CYP2A7* gene which was approximately 8 kb in size. The sequence comparison indicated that there were 13 nucleotide and 5 deduced amino acid differences between the coding region of the CoIIA and *CYP2A7* cDNA. These data suggest that the gene in CoIIA appears to be an allele of *CYP2A7*, and it is named *CYP2A7A*.

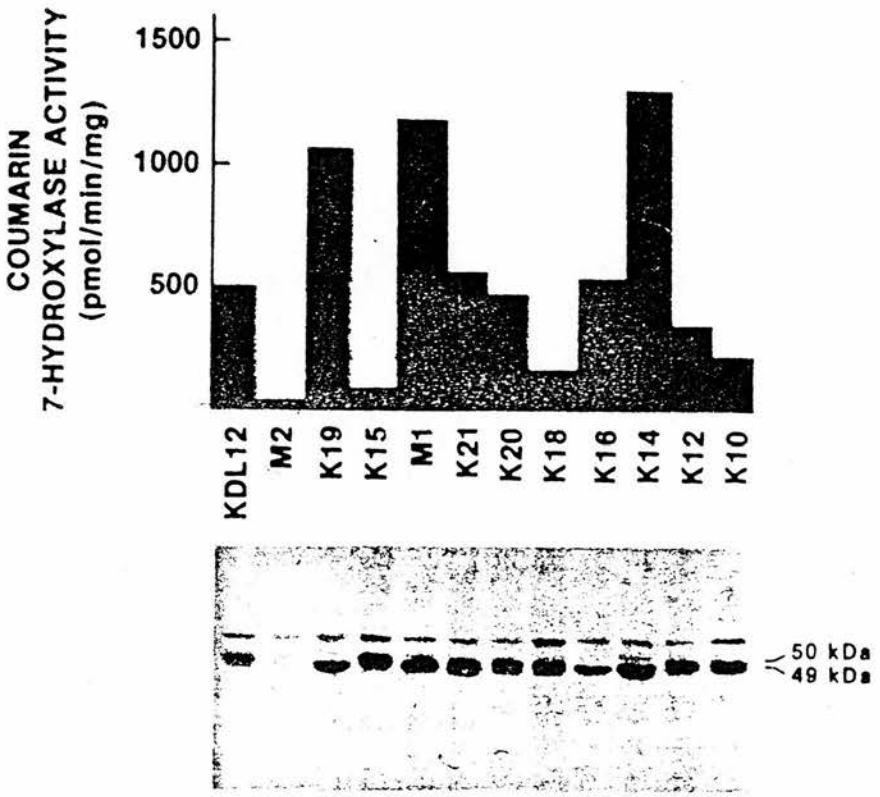
The second clone, LIIA, was isolated from a human genomic DNA library EMBL3. It did not hybridise with the *CYP2A7* exon 1 specific probe, but hybridised with the 3'-end specific probe (a fragment containing exon 5 to 9). A comparison of restriction maps and nucleotide sequences between the gene in clone LIIA and *CYP2A7A* indicates that the gene is not identical with *CYP2A7A*. The PCR/RFLP analysis of 18 human genomic DNA samples revealed that it appears to be another allele of *CYP2A7*, designated *CYP2A7B*. Of the 18 individuals tested, the frequency of *CYP2A7B* homozygotes is approximately 44.4%; that of *CYP2A7A* homozygotes is 33.3% and the frequency of heterozygotes of *CYP2A7A /CYP2A7B* is 22.2%. The primers used in PCR/RFLP are *CYP2A7* allele specific, and the PCR products contain no *CYP2A6* DNA. However, in the absence of the *CYP2A7* genomic sequence it is uncertain

whether the intronic amplification primers could actually amplify the *CYP2A7* gene. In addition, the restriction enzyme used in this analysis cannot distinguish the genotype of *CYP2A7A* from that of *CYP2A7*. Therefore, to isolate a full length *CYP2A7* gene in further studies will help us to determine the frequency of the *CYP2A7* genotype in humans.

5.2 The Relationship between the Alleles of *CYP2A7* and the Inter-individual Variability of Coh Activity in Man

It has been found that the level of Coh activity varies >100-fold in man (Yun *et al.*, 1991; Cholerton *et al.*, 1992). This is consistent with the difference in hepatic *CYP2A* gene expression determined by Northern blotting (Miles *et al.*, 1989a). The Western blot of human liver microsomal protein shows that the intensity of the stained 49 kDa protein band in the immunoblotting assay correlates with the level of the Coh activity in some tested microsomal samples. However, it is not difficult to notice that in some other samples, there is not such a correlation (Yamano *et al.*, 1990). In Figure 5.1, for example, both samples M2 and K18 have a low Coh activity, but the intensity of the immunoreactive 49 kDa protein band of the sample K18 is much higher than that of M2. My research results provided possible explanations of this lack of correlation at protein and mRNA levels. Firstly, the cDNA directed expression of *CYP2A6* and *CYP2A7* in COS-7 cells (Chapter 4) indicated that both cDNAs produced an identical 49 kDa protein, but only CYP2A6 had Coh activity. Secondly, it could be argued that although *CYP2A7* did not have Coh activity, it might not be expressed in most human livers, and therefore would not affect the correlation. However, the results of the expression levels of *CYP2A* genes in human liver demonstrated clearly that it was expressed in six tested liver samples, and that the ratio of CYP2A6 mRNA to CYP2A7 mRNA ranged between 1:0.5 to 1:3, respectively. These data suggest that the lack of correlation of the level of CYP2A6 protein with the Coh activity in some individuals, may be explained by the possibility that in these samples CYP2A7 is the major protein

Figure 5.1 Coh activities and Western immunoblotting analysis of CYP2A protein contents in 12 human liver specimens. Each liver microsome sample was subjected to immunoblotting analysis (20 μg/well) and coumarin 7-hydroxylase activity. The arrows denote the 50- and 49-kDa proteins, the latter of which corresponds to CYP2A (Yamano *et al.*, 1990).



present.

Two cDNAs, *CYP2A6* and *CYP2A7*, have been isolated from a human liver cDNA library, and an enzymatically inactive allele *CYP2A6v* has also been found in human liver (Yamano *et al.*, 1990). However, the phenotype study for coumarin 7-hydroxylation in man using thin-layer chromatography combined with fluorescence densitometry cannot explain the inter-individual variability on the basis of the alleles *CYP2A6* and *CYP2A6v* (Idle *et al.*, 1992).

In the studies on mouse Cyp2a, it has been found that three amino acid residues, Val117, Phe209 and Met365, play a very important role in coumarin 7-hydroxylation activity of Cyp2a5, and a single amino acid change (Val117 to Ala117) results in a 10-fold lower enzyme activity in some strains of mice (Lindberg *et al.*, 1989b; 1992; Gonzalez, 1992).

In humans, CYP2A6 is a Coh. The studies on the structure and regulation of *CYP2A7A* demonstrated that two genomic clones isolated from human genomic libraries were alleles of the *CYP2A7* gene, and this result was confirmed by PCR/RFLP analysis of human genomic DNAs (Chapter 3). Sequence comparison with *CYP2A6* which encoded a valine at position 117 showed that two base pair changes (GTA to GCG) in codon 117 of both *CYP2A7* and *CYP2A7A* resulted in a substitution of alanine for valine. However, *CYP2A7B* had only one silent base pair change (GTA to GTG) in codon 117, which did not lead to a change of the valine at this position (Table 5.1). This analysis suggests that the residue difference at position 117 among human *CYP2A* genes may be similar to *Cyp2a* genes, and hence responsible for the inter-individual variability in coumarin 7-hydroxylation in man. It is, therefore, important in future work to determine if there are any relationships between *CYP2A7A* or *CYP2A7B* homozygotes and the levels of the Coh activity. For example, *CYP2A7* allelic genotypes of individuals with high or low Coh activity could

Table 5.1 Scheme showing the relationships of important substitutions among the cDNAs of the *CYP2A* subfamily with their enzymatic activities. Coh represents coumarin 7-hydroxylase.

	117	209	365	Enzyme activity
Cyp2a-5Coh^{H}	Val	Phe	Met	High Coh activity
(mouse)				(Lindberg *et al.*, 1992)
Cyp2a-5Coh^{L}	Ala	Phe	Met	Low Coh activity
(mouse)				(Lindberg *et al.*, 1992)
Cyp2a4-15α	Ala	Leu	Leu	Steroid 15α-hydroxylase
(mouse)				(Lindberg *et al.*, 1992)
CYP2A3	Ala	Phe	Met	Low Coh activity ?
(rat)				(Gonzalez, 1992)
CYP2A6	Val	Phe	Val	Coh
(human)				(Miles *et al.*, 1990)
CYP2A7	Ala	Phe	Val	unknown
(human)				
CYP2A7A	Ala	Phe	Val	unknown
(human)				
CYP2A7B	Val			unknown
(human)				

be determined using allele-specific PCR in combination with the restriction enzyme *Mbo*I digestion described in Chapter 3. In view of the involvement of CYP2A enzymes in the metabolism of the tobacco-derived nitrosamine NNK and carcinogen aflatoxin B1, further studies on whether the genotypes of *CYP2A7* alleles are related with the metabolism of these procarcinognes will be helpful to understand the susceptibility of individuals to the toxic and carcinogenic effects of environmental chemicals.

5.3 Regulation of the *CYP2A7A* Gene

Some chemicals such as PB, DEX and TCPOBOP can induce an elevation in the transcription level of *CYP2A* genes in rodents (Pelkonen *et al.*, 1993) and in monkeys as well (Pearce *et al.*, 1992). However, the human *CYP2A* genes had not been extensively studied, because a suitable cell culture system is lacking for defining the inducible regulatory elements, and because no human *CYP2A* gene had ever been isolated and characterised. In this study, a human *CYP2A7A* gene was first isolated and then an attempt was made to determine the chemical inducible promoter activities of the gene in human HepG2 cells. The results revealed that the promoter activity of *CYP2A7A* was only slightly induced by DEX (50% increase) and pyrazole (61% increase). Phenobarbital treatment was inhibitory (58% decrease), TCPOBOP and ß-naphthoflavone were without effect on promoter activity. One explanation for these results is that although *CYP2A7A* is similar to the mouse *Cyp2a* genes in structure, regulatory elements involved in the expression are different in rodents and in humans. Another possible explanation is the absence of the specific transcription factors or receptors, which are responsible for the xenobiotic induction, in cultured HepG2 cells.

Results of the promoter activity assay described in Chapter 3 clearly showed that the 0.5 kb fragment of the 5'-flanking region of *CYP2A7A* gene revealed a maximal promoter activity. Progressive deletion of the HPF-1 binding site (5'-GCCAAAGTCCA-3') resulted in a 80% decrease in promoter activity. This result

shows that the HPF-1 binding site is an essential promoter element in basal transcription of human *CYP2A7A* in cultured HepG2 cells. Recently, Chen *et al.* (1994) reported that the HPF-1 binding site was highly homologous to the hepatocyte nuclear factor 4 (HNF-4) binding motif, and both factors (HPF-1 and HNF-4) can bind to either HPF-1 or HNF-4 binding site. These findings suggest that HPF-1 might be an important factor in the hepatic expression of *CYP2A7A*. In future studies, other approaches can be adopted to demonstrate the mechanism of the constitutive and tissue specific expression of P450s. For example, mobility shift of DNA-binding assay can be carried out to determine the binding reaction between regulatory proteins, such as HNF-4 factor, and promoter elements of *CYP2A7A* using protein extracts from liver or from other tissues. This technique can also be used to characterise the mechanism of liver specific expression of P450. Co-transfection of the construct containing the promoter element of the *CYP2A7A* gene with another construct, which is capable of generating HNF-4 protein, into COS cells could be an alternative way to determine the function of the HPF-1 motif in liver-specific expression.

It has been found that the steroid regulatory element (SRE, 5'-CACCCCAC-3') plays an essential role in co-ordinating the transcriptional regulation of the genes involved in the maintenance of cholesterol homeostasis, and functions as an enhancer (Osborne *et al.*, 1988; Smith *et al.*, 1988). A SRE has also been reported existing in the promoter region of the human *CYP7* gene which is responsible for the metabolism of cholesterol and bile acid biosynthesis (Molowa *et al.*, 1992). In this study a consensus SRE was identified by sequencing the 0.5 kb 5' flanking region of *CYP2A7A*, and the SRE was found overlapping with two directly repeated CACCC elements (CCACCCCACCC), which function like Sp1 binding site. This is similar to the promoter region in the cholesterol homeostasis related LDL receptor gene, in which the SRE is located between two Sp1 binding sites (Goldstein and Brown, 1990). Although the function of the *CYP2A7A* gene is still unclear, this finding together with the fact that mouse Cyp2a proteins metabolise steroid hormones (Lindberg *et al.*, 1989a; Burkhart *et al.*, 1990)

suggest that the function and regulation of the *CYP2A7A* gene may be related to steroid metabolism.

It is interesting to notice that by the transient transfection assay, the construct containing a 3.0 kb 5'-flanking region of *CYP2A7A* exhibited only a low promoter activity in HepG2 cells. With the progressive deletion of the 3.0 kb 5'-flanking sequence to the position of -1.0 kb, the transcription activity significantly increased and the maximal activity was observed with a fragment containing approximately 0.5 kb of the 5' flanking region of the gene (Chapter 3). This result suggests that in human liver, the expression of *CYP2A7A* is under multiple control mechanisms, possibly including positive and negative control elements.

In order to characterise these control elements, a series of fused constructs could be made in future studies, which contain a reporter and a truncated fragment from different parts of the 3.0 kb 5'-flanking region of *CYP2A7A*. These constructs can be transfected into cultured cells to determine their transcriptional activities. This study will shed light on why the 3.0 kb 5'-flanking region of the *CYP2A7A* only has a low promoter activity in HepG2 cells.

5.4 Alternative Splicing of CYP2A7

As part of the analysis of CYP2A6 and CYP2A7 mRNAs, RT-PCR was carried out to amplify full length *CYP2A* cDNAs using a human liver RNA sample L8 (Chapter 4). The results showed that in addition to the expected fragment of 1.6 kb, a small amount of a 1.45 kb DNA fragment was observed after electrophoresis separation on an agarose gel. The PCR products were subcloned into pUC19, and a restriction enzyme *Pst*I digestion was used to screen the colonies as a *Pst*I site was only present at 143 bp (+1 indicates the start of the open reading frame) of *CYP2A6* cDNA. Three clones with different digestion patterns were isolated and sequenced. The first two, both with a 1.6 kb insert, contained a *CYP2A6* and a *CYP2A7* cDNA, respectively. The third one,

with a 1.45 kb insert, contained an alternatively spliced version of *CYP2A7* (*CYP2A7AS*).

Comparing the sequence of *CYP2A7AS* cDNA with the genomic DNA of *CYP2A7*, it was found that both coding regions were identical except that the 163 bp exon 2 of *CYP2A7AS* was replaced by a 10 bp segment of intron 1. Translation of CYP2A7AS mRNA resulted in an in-frame deletion of 51 amino acids, and in the generation of a protein product of M_r 44 kDa. The 10 bp segment of intron 1 added three amino acids at residue 60 and was then linked to the amino acid 114 in exon 3. All the intron/exon junctions conformed to the GT/AG consensus splice recognition site. However, intron 1 contained an additional splice site, G/gcagg (exon sequence is designated by upper-case letter and intron sequences are designated by lower-case letters), which resulted in CYP2A7AS. Interestingly, CYP2A7AS was the major CYP2A7 mRNA detected in a human skin fibroblast cell line. This finding shows that the alternative splicing of CYP2A7 may be tissue-specific.

The cDNA directed expression of the *CYP2A6*, *CYP2A7* and *CYP2A7AS* in COS-7 cells described in this study indicate that both CYP2A6 and CYP2A7 products are 49 kDa, and CYP2A7AS is a truncated protein of *M*r 44 kDa. But only CYP2A6 has a coumarin 7-hydroxylation activity. To date, no substrate for CYP2A7 has been identified. The analyses of CYP2A mRNA levels in human livers clearly demonstrated that in certain liver samples *CYP2A7* was expressed at higher levels than *CYP2A6* , or even was a dominant mRNA species of CYP2A. It will, therefore, be important to determine the substrate(s) for this enzyme in future studies. In addition, as the expression system used for P450 expression in mammalian cells is usually not efficient enough to characterise protein biosynthesis, structure and function of P450, further studies on other high expression systems, such as amplifiable systems leading to an overproduction of foreign proteins in mammalian cells, will provide evidence for the characterisation of the functions of CYP2A7A. Alternatively, a cell line stably

expressing *CYP2A7* cDNA can be established to determine the function of CYP2A7 in chemical metabolism, and to evaluate its effects on the toxic and carcinogenic potency of a chemical.

BIBLIOGRAPHY

Ali, I.U., Reinhold, W., Salvador, C. and Aguanno, S. (1992) Aberrant splicing of Gsa transcript in transformed human astroglial and glioblastoma cell lines. *Nucleic Acids Res.*, **20**, 4263-4267.

Albano, E., Tomasi, A., Persson, J.O., Terelius, Y., Goria-Gatti, L., Ingelman-Sundberg, M. and Dianzani, M.U. (1991) Role of ethanol-inducible cytochrome P-450-450 (P450IIE1) in catalysing the free radical activation of aliphatic alcohols. *Biochem. Pharmacol.*, **41**, 1895-1902.

Alvan, G., Bechtel, P, Iselius, L. and Gundert-Remy, U. (1990) Hydroxylation polymorphisms of debrisoquine and mephenytoin in European populations. *Eur. J. Clin. Pharmacol.*, **39**, 533-537.

Andersson, S., Davis, D.L., Dahlbäck, H., Jörnvall, H. and Russell, D.W. (1989) Cloning, structure, and expression of the mitochondrial cytochrome P-450 sterol 26-hydroxylase, a bile acid biosynthetic enzyme. *J. Biol. Chem.*, **264**, 8222-8229.

Aoyama, T., Yamano, S., Guzelian, P.S., Gelboin, H.V. and Gonzalez, F.J. (1990) Five of 12 forms of vaccinia virus-expressed human hepatic cytochrome P450 metabolically activate aflatoxinB1. *Proc. Natl. Acad. Sci. USA*, **87**, 4790-4793.

Arinc, E. (1993) Extrahepatic Microsomal forms: Lung microsomal cytochrome P450 isozymes. In Schenkman, J.B. and Greim, H. (eds) Cytochrome P450, Handbook of Experimental Pharmacology, Vol. **105**, Springer-Verlag, 373-385.

Ashby, J. and Tennant, R.W. (1991) Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutation Res.*, **257**, 229-306.

Atchison, M. and Adesnik, M. (1986) Gene conversion in a cytochrome P-450 gene family. *Proc. Natl. Acad. Sci. USA*, **83**, 2300-2304.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987) Current Protocols in Molecular Biology. Greene Publishing Associates and Wiley-Interscience.

Bar-Nun, S., Kreibich, G., Adesnik, M., Alterman, L., Negishi, M. and Sabatini, D.D. (1980) Synthesis and insertion of cytochrome P-450 into endoplasmic reticulum membrane. *Proc. Natl. Acad. Sci. USA,* **77**, 965-969.

Bascom, R. (1991) Differential susceptibility to tobacco smoke: possible mechanisms. *Pharmacogenetics*, **1**, 102-106.

Baumhueter, S., Coutois, C and Crabtree, G.R. (1988) A variant nuclear protein in dedifferentiated hepatoma cells binds to the same functional sequences in the ß fibrinogen gene promoter as HNF-1. *EMBO J.*, **7**, 2485-2493.

Baumhueter, S., Mendel, D.B., Conley, P.B. Kuo, C.J. Tuek, C., Graves, M.K., Edwards, C.A., Courtois, G. and Crabtree, G.R. (1990) HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-BI and APF. *Gene & Dev.*, **4**, 372-379.

Bergh, A.F. and Strobel, H.W. (1992) Reconstitution of brain mixed function oxidases system: purification of NADPH-cytochrome P450 reductase and partial purification of cytochrome P450 from whole rat brain. *J. Neurodrem*, **59**, 575-581.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant Plasmid DNA. *Nucleic Acids Res.*, **7**, 1513-1518.

Black, S.D. (1992) Membrane topology of the mammalian P450 cytochromes. *The FASEB J.*, **6**, 680-685.

Black, S.D. (1993) Cytochrome P450 structure and function. In: Schenkman, J.B. and Greim, H. (eds) Cytochrome P450, Handbook of Experimental Pharmacology, Vol. **105**, Springer-Verlag, 155-168.

Blake, R.C. II and Coon, M.J. (1989) On the mechanism of action of cytochrome P-450. Spectral intermediates in the reaction with iodosobenzene and its derivatives. *J. Biol. Chem.*, **264**, 3694-3701.

Blanck, J, Ristau, O., Zhukov, A.A., Archakov, A.I., Rein, H. and Ruckpaul, K. (1991) Cytochrome P-450 spin state and leakiness, of the monooxygenase pathway. *Xenobiotica*, **21**, 121-135.

Boobis, A.R., Lynch, A.M., Murray, S., Torre, R.de la, Solans, A., Farré, M., Segura, J., Gooderham, N.J. and Davies, D. (1994) CYP1A1-catalyzed conversion of

dietary heterocyclic amines to their proximate carcinogens is their major route of metabolism in humans. *Cancer Research* , **54**, 89-94.

Broly, F., Gaedigk, A., Heim, M., Eichelbaum, M., Morike, K. and Meyer, U.A. (1991) Debrisoquine/sparteine hydroxylation genotype and phenotype: analysis of common mutations and alleles of CYP2D6 in a European population. *DNA Cell Biol.*, **10**, 545-501.

Brosen, K. and Gram, L.F. (1989) Clinical significance of the sparteine/debrisoquine oxidation polymorphism. *Eur. J. Clin. Pharmacol.*, **36**, 537-547.

Burkhart, B.A., Skow, L.C. and Negishi, M. (1990) Two steroid 15α-hydroxylase genes and a homologous gene family in mice. *Gene*, **87**, 205-211.

Butler, M.A., Lang, N.P., Young, J.F., Caporaso, N.E., Vinesi, P., Hayes, R.B., Teitel, C.H., Massengill, J.P., Lawson, M.F. and Kadlubar, F.F. (1992) Determination of CYP1A2 and N-acetyltransferase 2 phenotypes in human populations by analysis of caffeine urinary metabolites. *Pharmacogenetics*, **2**, 116-127.

Camus, A.M., Geneste, O., Honkakoski, P., Bereziat, J.-C., Henderson, C.J., Wolf, C.R. and Lang, M.A., *Carcinogenesis*, (in press).

Chae, Y-H., Yun, C-H, Guengerich, F.P. Kadlubar, F.F. and EL-Batoumy, K. (1993) Roles of human hepatic and pulmonary cytochrome P450 enzymes in the metabolism of the environmental carcinogen 6-Nitrochrysene. *Cancer Res.*, **53**, 2028-2034.

Chan, G. and Kemper, B. (1990) Structure of the rabbit P450IIC3 gene, a constitutive member of the P450IIC subfamily. *Biochemistry*, **29**, 3743-3750.

Chen, C. and Okayama, H. (1987) High-efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell Biol.*, **7**, 2745-2752.

Chen, D., Lepar, G. and Kemer, B. (1994) A transcriptional regulatory element common to a large family of hepatic cytochrome P450 genes is a functional binding site of the orphan receptor HNF-4. *J. Biol. Chem.*, **269**, 5420-5427.

Chen, T.R. (1987) *In situ* detection of mycoplasma contamination in cell cultures by fluorescent Hoescht 33258 stain. *Exp. Cell Res.*, **104**, 255-262.

Cholerton, S., Idle, M.E., Vas, A., Gonzalez, F.J. and Idle, J.R. (1992) Comparison of a novel thin-layer chromatographic-fluorescence detection method with a spectrofluorometric method for the determination of 7-hydroxycoumarin in human urine. *J. Chromatogr.*, **575**, 325-330

Chomczynski, O. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate. *Anal. Biochem.*, **162**, 156-159.

Clark, L. and Waxman D.J. (1989) Oxidative metabolism of cyclophosphamide: identification of the hepatic monooxygenase catalysts of drug activation. *Cancer Res.*, **49**, 2344-2350.

Costa, R.H., Grayson, D.R. and Darnell, J.E. (1989) Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and 1-antitrypsin genes. *Mol. Cell. Biol.*, **9**, 1415-1425.

Crespi, C.L., Penman, B.W., Leakey, J.A.E., Arlotto, M.P., Stark, A., Parkinson, A., Turner, T., Steimed, D.T., Rudo, K., Davies, R.L. and Langenbach, R. (1990) Human cytochrome P450IIA3: cDNA sequence, role of the enzyme in the metabolic activation of promutagens, comparison to nitrosamine activation by human cytochrome P450IIE. *Carcinogenesis*, **11**, 1293-1300.

Crespi, C.L., Penman, B.W., Steimel, D.T., Gelboin, H.V. and Gonzalez, F.J. (1991a) The development of a human cell line stably expressing human CYP3A4: role in the metabolic activation of aflatoxin B_1 and comparison to CYP1A2 and CYP2A3. *Carcinogenesis*, **12**, 355-359.

Crespi, C.L., Penman, B.W., Gelboin, H.V. and Gonzalez, F.J. (1991b) A tobacco-smoke derived nitrosamine, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone, is activated by multiple human cytochrome P450s including the polymorphic cytochrome P4502D6. *Carcinogenesis*, **12**, 1197-1201.

Cullen, B.R. (1986) Use of eukaryotic expression technology in the functional analysis of cloned genes. In: Methods in Enzymology, Academic Press, Inc., Vol. **152**, 684-704.

Curnow, K.M., Slutsker, L., Vitek, J., Cole, T., Speiser, P.W., New, M.I., White, P.C. and Pascoe, L. (1993) Mutations in the CYP11B1 gene causing congenital adrenal hyperplasia and hypertension cluster in exon 6, 7, and 8. *Proc. Natl. Acad. Sci. USA*, **90**, 4552-4556.

Daly, A.K., Armstrong, M., Monkman, S.C., Idle, M.E. and Idle, J.R. (1991) The genetic and metabolic criteria for the assignment of debrisoquine hydroxylation (cytochrome P450IID6) phenotypes. *Pharmacogenetics*, **1**, 33-41.

Daly, A.K. and Idle, J.R. (1993) Genetics: animal and human cytochrome P450 polymorphisms. In Schenkman, J.B. and Greim, H. (eds) Cytochrome P450, Handbook of Experimental Pharmacology, Springer-Verlag, Vol. **105**, 433-446.

Davies, R.L., Crespi, C.L., Rudo, K., Turner, T.R. and Langenbach, R. (1989) Development of a human cell line by selection and drug-metabolizing gene transfection with increased capacity to activate promutagens. *Carcinogenesis*, **10**, 885-891.

Davis, G.C., Willian, A.C., Markey, S.P., Ebert, M.H., Caine, E.D., Reichert, C.M., Kopin, I.J. (1979) Chronic Parkinsonisms secondary to intravenous injection of meperidine analogues. *Psychiatry Res.* **1**, 249-254.

Doehmer, J., Dogra, S., Friedberg, T., Monier, S., Adesnik, M., Glatt, H. and Oesch, F. (1988) Stable expression of rat cytochrome P-450IIB cDNA in Chinese hamster cell (V79) and metabolic activation of aflatoxin B_1. *Proc. Natl. Acad. Sci. USA*, **85**, 5769-5773.

Druckrey, H., Preussmann, R. and Ivankovic, S. (1969) N-Nitroso compounds in organotropic and transplacental carcinogenesis. *Ann. NY Acad. Sci.*, **163**, 676-696.

Eichelbaum, M., Kroemer, K.H. and Mikus, G. (1992a) Genetically determined differences in drug metabolism as a risk factor in drug toxicity. *Toxicology letters*, **1**, 115-122.

Eichelbaum, M. and Gross, A.S. (1992b) The genetic polymorphism of debrisoquine/sparteine metabolism-clinical aspects. In: Kalow, W. (ed.), Pharmacogenetcs of Drug metabolism. Peramon Press, New York, 625-648.

English, N., Hughes, V. and Wolf, C.R. (1994) Peroxisome proliferators mimic an endogenous inducer and inactivate a transcriptional repressor in *Bacillus megaterium.* In Cockburn, A. and Smith, L. (eds) Nongenotoxic Carcinogenesis, Ernst Schering Research Foundation Workshop 10, Springer-Verlag, 201-219.

Estabrook, R.W., Cooper, D.Y., Rosenthal, O. (1963) The light reversible carbon monoxide inhibition of the steroid C21-hydroxylase system of the adernal cortex. *Biochem. Z.*, **338**, 741-755.

Feinberg, D.P. and Vogelstein, B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **136**, 6-13.

Fonne-Pfister, R., Bargetzi, M.J.A., Meyer, U.A. (1987) MPTP the neurotoxin inducing Parkinson's disease, is a potent inhibitor of human and rat P450 enzymes (P450bufl, P450db1) catalyzing debrisoquine 4-hydroxylation. *Biochem. Biophys. Res. Commun.*, **148**, 1144-1150.

Forrester, L. M., Neal, G.E., Judah, D.J., Glancey, M.J. and Wolf, C.R. (1990) Evidence for involvement of multiple forms of cytochrome P-450 in aflatoxin B_1 metabolism in human liver. *Proc. Natl. Acad. Sci. USA,* **87**, 8306-8310.

Forrester, L.M., Henderson, C.J., Glancey, M.J., Back, D.J., Park, B.K., Ball, S.E., Kitteringham, N.R., Mclaren, A.W., Miles, J.S., Skett, P. and Wolf, C.R. (1992) Relative expression of cytochrome P450 isoenzymes in human liver and association with the metabolism of drugs and xenobiotic. *Biochem. J.*, **281**, 359-368.

Frain, M., Swart, G., Monaci, P., Nicosia, A., Stampfli, S., Frank, R. and Cortese, R. (1989) The liver-specific transcriptional factor LF-BI contains a highly diverged homeobox DNA binding domain. *Cell*, **59**, 145-157.

Freshney, R.I. (1987) Culture of animal cells: A manual of basic technique. Liss, A.R., Inc.

Fukuhara, M., Nagata, K., Mizokami, K., Yamazoe, Y., Takanaka, A. and Kato, R. (1989) Complete cDNA sequence of a major 3-methylcholanthrene-inducible cytochrome P-450 isozyme (P450AFB) of syrian hamsters with high activity toward aflatoxin B_1. *Biochem. Biophys. Res. Commun.*, **162**, 265-272.

Gibson, G.G. and Skett, P. (1986) Introduction to drug metabolism. Chapman and Hall Ltd, 1-38.

Goldstein, J. and Brown, M. (1990) Regulation of the mevalonate pathway. *Nature*, **343**, 425-430.

Gonzalez, F.J., Skoda, R.C., Kimula, S., Umeno, M., Zanger, U.M., Nebert, D.W., Gelboin, H.V. Hardwick, J.P. and Meyer,U.A. (1988) Characterisation of the common genetic defect in humans deficient in debrisoquine metabolism. *Nature*, **331**, 442-446.

Gonzalez, F.J. (1989) The molecular biology of cytochrome P450s. *Pharmacol. Rev.*, **40**, 243-287.

Gonzalez, F.J. (1990) Molecular genetics of the P-450 superfamily. *Pharmac. Ther.*, **45**, 1-38.

Gonzalez, F.J. and Nebert, D.W. (1990) Evolution of the P450 gene superfamily: animal-plant "warfare", molecular drive, and human genetic differences in drug oxidation. *Trends in Genetics*, **6**, 182-186.

Gonzalez, F.J. (1992) Human cytochromes P450: problems and prospects. *TiPS,* **13**, 346-352.

Gonzalez, F.J. (1993) Cytochrome P450 evolution and nomenclature. In Schenkman, J.B. and Greim, H. (eds) Cytochrome P450, Handbook of Experimental Pharmacology, Springer-Verlag, Vol. **105**, 211-220.

Gonzalez, F.J., Liu, Su-Yan and Yano, M. (1993) Regulation of cytochrome P450 genes: molecular mechanisms. *Pharmacogenetics*, **3**, 51-57.

Gorsky, L.D., Koop, D.R. and Coon, M.J. (1984) On the stoichiometry of the oxidase and monooxygenase reaction catalysed by liver microsomal cytochrome P450: Product of oxygen reduction. *J. Biol. Chem.*, **259**, 6812-6817.

Gotoh, O. and Fujii-kuriyama, Y. (1989) Evolution, structure and gene regulation of cytochrome P-450. In: Ruckpaul, K. and Rein, H. (eds), Frontiers in biotransformation: I. Basis and mechanisms of regulation of cytochrome P-450. Taylor and Francis, London, 195-243.

Gotoh, O. (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. **267**, *J. Biol. Chem.*, **267**, 83-90.

Gotoh, O. (1993) Evolution and differentiation of P-450 genes. In Omura, T., Ishimura, Y. and Fujii-Kuriyama, Y. (eds) Cytochrome P-450, Kodansha Ltd. and VCH Publishers Inc., 255-272.

Gough, A.C., Miles, J.S., Spurr, N.K., Moss, J.E., Gaedigk, A., Eichelbaum, M. and Wolf, C.R. (1990) Identification of the primary gene defect at the cytochrome P450 CYP2D locus. *Nature*, **347**, 773-776.

Groopman, J.D. and Donahue, K.F. (1988) Aflatoxin, a human carcinogen: determination in foods and biological samples by monoclonal antibody affinity chromatography. *J. Assoc. Anal. Chem.*, **71** (special rep.), 861-867.

Guengerich, F.P. (1988) Roles of cytochrome P-450 enzymes in chemical carcinogenesis and cancer chemotherapy. *Cancer Res.*, **48**, 2946-2954.

Guengerich, F.P. (1991) Reactions and significance of cytochrome P-450 enzymes. *J. Biol. Chem.*, **266**, 10019-10022.

Guo, W. and Helfman, D.M. (1993) *Cis*-elements involved in alternative splicing in the rat β-tropomyosin gene: the 3' splice site of the skeletal muscle exon 7 is the major site of blockage in nonmuscle cells. *Nucleic Acids Res.*, **21**, 4762-4768.

Hagen, G., Müller, S., Beato, M. and Suske, G. (1992) Cloning by recognition site screening of two novel GT box binding proteins: a family of Sp1 related genes. *Nucleic Acids Res.*, **20**, 5519-5525.

Hahn, C., Hansen, A.J. and May, B.K. (1991) Transcriptional regulation of the chicken CYP2H1 gene. *J. Biol. Chem.*, **266**, 17031-17039.

Hahnemann, B., Salonpää, P., Pasanen, M., Mäenpää, J., Honkakoski, P., Juvoene, R., Lang, M.A., Pelkoene, O. and Raunio, H. (1992) Effect of pyrazole, cobalt and phenobarbital on mouse liver cytochrome P-450 2a-4/5 (Cyp2a-4/5) expression. *Biochem. J.*, **286**, 289-294.

Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* , **166**, 557-580.

Hardwick, J.P., Gonzalez, F.L. and Kasper, C.B. (1983) Transcriptional regulation of rat liver epoxide hydratase, NADPH-cytochrome P-450 oxidoreductase, and cytochrome P-450b genes by phenobarbital. *J. Biol. Chem.*, **258**, 8081-8085.

Harris, C.C. (1991) Chemical and physical carcinogenesis: advances and perspectives for the 1990s. *Cancer Res.*, **51**, 5023-5044.

Hayashi, D. and Nozaki, M. (1969) Nature and mechanisms of oxygenase. *Science*, **164**, 389-3396.

He, J-S. and Fulco, A. (1991) A barbiturate-related protein binding to a common sequence in the cytochrome P450 genes of rodents and bacteria. *J. Biol. Chem.*, **266**, 7864-7869.

Hecht, S.S., Castonguay, A., Rivenson, A., Mu, B. and Hoffmann, D. (1983) Tobacco specific nitrosamines: carcinogenicity, metabolism and possible role in human cancer. *J. Environ. Sci. Health Part C*, **1**, 1-54.

Hecht, S.S. and Hoffmann, D. (1988) Tobacco-specific notrosamines, an important group of carcinogens in tobacco and tobacco smoke. *Carcinogenesis*, **9**, 875-884.

Henderson, C.J. and Wolf, C.R. (1992) Molecular analysis of cytochrome P450s in the CYP2 gene family. In Gibson, G.G., Progress in Drug Metabolism. Taylor & Francis Ltd. Vol. **13**, pp 73-87.

Honkakoski, P., Kojo, A. and Lang, M.A. (1992) Regulation of the mouse liver cytochrome P450 2B subfamily by sex hormones and phenobarbital. *Biochem. J.*, **285**, 979-983.

Hotchkiss, S.A.M. (1992) Skin as a xenobiotic metabolizing organ. In Gibson, G.G., (eds), Progress in drug metabolism, Taylor & Francis Ltd. Vol. **13**, 217-262.

Idle, J.R., Armstrong, M., Boddy, A.V., Boustread, C., Cholerton, S., Cooper, J., Daly, A.K., Ellis, J., Gregory, W., Hadidi, H., Höfer, C., Holt, J., Leathart, J., McCracken, N., Monkman, S.C., Painter, J.E., Taber, H., Walker, D. and Yule, M. (1992) The pharmacogenetics of chemical carcinogenesis. *Pharmacogenetics*, **2**, 246-258.

Iersel, M.L.P.S. van, Henderson, C.J., Walter, D.G., Price, R.J., Wolf, C.R. and Lake, B.G. (in press) Metabolism of (3-^{14}C) coumarin by human liver microsomes.

Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (1990) PCR Protocols. A Guide to Methods and Applications. Academic Press Inc., San Diego, California, 21-27.

Ish-Horowicz, D. and Burke, J.F. (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Res.*, **9**, 2989.

Jaiswal, A.K., Haaparanta, T., Luc, P-V., Schembri, J. and Adesnik, M. (1990) Glucocorticoid regulation of a phenobarbital-inducible cytochrome P-450 gene: the

presence of a functional glucocorticoid response element in the 5'-flanking region of the CYP2B2 gene. *Nucleic Acids Res.*, **18**, 4237-4242.

Jantzen, H-M., Strähle, U., Gloss, B., Stewart, F., Schmid, W., Boshart, M., Miksicek, R. and Schütz, G. (1987) Cooperativity of glucocorticoid response elements located far upstream of the tyrosine aminotransferase gene. *Cell*, **49**, 29-38.

Johnson, P.F., Landschulz, W.H., Graves, B.J. and Mcknight, S.L. (1987) Identification of a rat liver nuçlear protein that binds to the enhancer core element of three animal viruses. *Gene & Dev.*, **1**, 133-146.

Johansson, I., Lundqvist, E., Bertilsson, L., Dahl, M-L., Sjöqvist, F. and Ingelman-Sundberg, M. (1993) Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. *Proc. Natl. Acad. Sci. USA*, **90**, 11825-11829.

Jung, C. and Ristau, O. (1978) Mechanism of the cytochrome P-450 catalyzed hydroxylation-thermodynamic aspects and the nature of the active oxygen species. *Pharmazie,* **33**, 329-331.

Kagimoto, M., Heim, M., Kagimoto, K., Zeugin, T. and Meyer, U.A. (1990) Multiple mutations of the human cytochrome P450IID6 gene (CYP2D6) in poor metabolizers of debrisoquine. *J. Biol. Chem.*, **265**, 17209-17214.

Kappus, H. (1993) Metabolic reactions: role of cytochrome P-450 in the formation of reactive oxygen species. In Schenkman, J.B. and Greim, H., Cytochrome P450, Springer-Verlag, Berlin Heidelberg, 147-154.

Kato, R., Yasumori, T. and Yamazoe, Y. (1991) Characterization of human P450IIC isozymes by using a yeast expression system. In Waterman, M.R. and Johnson, E.F. (eds.) Methods in Enzymology: Cytochrome P450. Academic Press, Inc., Vol. **206**, 183-190, .

Kawajiri, K., Nakachi, K., Imai, K., Yoshii, A., Shinoda, N. and Watanabe, J. (1990) Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P4501A1 gene. *FEBS Lett.*, **263**, 131-133.

Keyse, S.M. & Emslie, E.A. (1992) Oxidative stress and heat shock induce a human gene encoding a protein-tyrosine phosphatase. *Nature*, **359**, 644-647.

Khandjian, E.W. (1987) Optimized hybridisation of DNA blotted and fixed to nitrocellulose and nylon membranes. *Bio/Technology,* **5**, 165-168.

Kimura, S., Kozak, C.A. and Gonzalez, F.J. (1989) Identification of a novel P450 expressed in rat lung: cDNA cloning and sequence, chromosome mapping, and induction by 3-methylcholanthrene. *Biochemistry*, **28**, 3798-3803.

Kimura, H., Sogawa, K., Sakai, Y. and Fujii-Kuriyama, Y. (1989) Alternative splicing mechanism in a cytochrome P-450 (P-450PB-1) gene generates the two mRNAs coding for proteins of different functions. *J. Biol. Chem.*, **264**, 2338-2342.

Kioussis, D., Wilson, F., Daniels, C., Leveton, C., Taverne, J. and Playfair, J.H.L. (1987) Expression and rescuing of a cloned human tumour necrosis factor gene using an EBV-based shuttle cosmid vector. *EMBO J.*, **6**, 355-361.

Klingenberg, M. (1958) Pigments of rat liver microsomes. *Arch. Biochem. Biophys.* **75**, 376-386.

Knowles, B.B., Howe, C.C. and Aden, D.P. (1980) Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. *Science*, **209**, 497-499.

Koga, N., Ariyoshi, N., Nakashima, H. and Yoshimura, H. (1990) Purification and characterization of two forms of 2,3,4,7,8-pentachlorodibenzofuran-inducible cytochrome P-450 in hamster liver. *J. Biochem.* (Tokyo), **107**, 826-833.

Kominami, S. (1993) Substrate binding and the reduction of cytochrome P-450. In Omura, T., Ishimura, Y. and Fujii-Kuriyama, Y. (eds) Cytochrome P-450, Kodansha Ltd. and VCH Publishers Inc., 64-80.

Kosaki, A. and Webster, N.J.G. (1993) Effect of dexamethasone on the alternative splicing of the insulin receptor mRNA and insulin action in HepG2 hepatoma cells. *J. Biol. Chem.*, **268**, 21990-21996.

Kouri, R.E., Mckinney, C.E., Slomiany, D.J., Snodgrass, D.R., Wray, N.P.and Mclemore, T.L. (1982) Positive correlation between high aromatic hydrocarbon hydroxylase activity and primary lung cancer as analysed in cryopreserved lymphocytes. *Cancer Res.*, **42**, 5030-5037.

Kuo, C.J., Conley, P.B., Chen, L., Sladek, F.M., Darnell, J.E., Jr (1992) A transcriptional hierarchy involved in mammalian cell-type specification. *Nature*, **355**, 457-461.

Kuthan, H. and Ullrich, V. (1982) Oxidase and oxygenase function of the microsomal cytochrome P-450 monooxygenase system. *Eur. J. Biochem.*, **126**, 583-588.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-685.

Lai, E. and Darnell, J.E., Jr (1991) Transcriptional control in hepatocytes: a window on development. *Trends Biochem. Sci.*, **16**, 427-430.

Lai, T.S. and Chiang, J.Y.L. (1990) Cloning and characterisation of two major 3-methylcholanthrene inducible hamster liver cytochrome P-450s. *Arch. Biochem. Biophys.*, **283**, 429-439.

Lange, R., Périn, F., Larroque, C. and Zajdela, F. (1990) Isolation and partial characterisation of a cytochrome P-450 isoenzyme (cytochrome P-450_{tu}) from mouse liver tumors. *Biochim. Biophys. Acta*, **1038**, 130-135.

Larrey, D., Distelrath, L.M., Dannan, G.A., Wilkinson, G., Guengerich, F.P. (1984) Purification and characterisation of the rat liver microsomal cytochrome P450 involved in the 4-hydroxylation of debrisoquine, a prototype for genetic variation in oxidative drug metabolism. *Biochemistry*, **23**, 2787-2795.

Lewis, A.D., Hickson, T.D., Robson, C.N., Harris, A.L., Hayes, J.D., Griffiths, S.A., Manson, M.M., Hall, A.E., Moss, J.E. and Wolf, C.R. (1988) Amplification and increased expression of alpha class glutathione S-transferase encoding genes associated with resistance to nitrogen mustards. *Proc. Natl. Acad. Sci. USA*, **85**, 8511-8515.

Li, W-H. and Graur, D. (1991) Evolution by gene duplication and exon shuffling, in Fundamentals of molecular evolution. Sinauer Associates, Inc. Publishers, 136-171.

Lin, L., Young, F., Ye, Z., Xu, E., Young, C., Zhang, C.,Wu, D. and Nebert, D.W. (1991) Case-control study of cigarette smoking and primary hepatoma in an aflatoxin-endemic region of China: a protective effect. *Pharmacogenetics*, **1**, 79-85.

Lindberg, R.L.P., Burkhart, B., Ichikawa, T. and Negishi, M. (1989a) The structure and characterization of type I P-$450_{15\alpha}$ gene as major steroid 15α-hydroxylase and its comparison with type II P-$450_{15\alpha}$. *J. Biol. Chem.*, **264**, 6465-6471.

Lindberg, R.L.P. and Negishi, M. (1989b) Alteration of mouse cytochrome $P450_{coh}$ substrate specificity by mutation of a single amino-acid residue. *Nature*, **339**, 632-634.

Lindberg, R.L.P., Juvonen, R. and Negishi, M. (1992) Molecular characterization of the murine Coh locus: an amino acid difference at position 117 confers high and low coumarin 7-hydroxylase activity in $P450_{coh}$. *Pharmacogenetics*, **2**, 32-37.

Lowry, D.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951) Protein measurement with folin phenol reagent. *J. Biol. Chem.*, **193**, 265-275.

Lu, A.Y.H. and Coon, M.J. (1968) Role of hemoprotein P450 in fatty acid ω-hydroxylation in a soluble enzyme from liver microsomes. *J. Biol. Chem.*, **243**, 1331-1332.

Magee, P.N. and Barnes, J.M. (1967) Carcinogenic nitroso compounds. *Adv. Cancer Res.*, **10**, 163-246.

Mahgoub, A., Idle, J.R., Dring, L.G., Lancaster, R. and Smith, R.L. (1977) Polymorphic hydroxylation of debrisoquine in man. *Lancet*, **2**, 584-586.

Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, D.Q., Sim, G.K. and Efstratiadis, A. (1978) The isolation of structural genes from libraries of eukaryotic DNA. *Cell,* **15**, 687.

Maniatis, T, Fritsch, E.F and Sambrook, J. (1989) Molecular Cloning. A Laboratory Manual. Second edition. Cold Spring Harbor University Press. Cold Spring Harbor.

Matsunaga, T., Nagata, K., Holsztynska, E.V., Lapenson, D.P. Smith, A.J., Kato, R., Gelosin, H.V., Waxman, D.J. and Gonazalez, F.J. (1988) Gene conversion and differential regulation in the rat P-450IIA gene subfamily: purification, catalytic activity, cDNA and deduced amino acid sequence, and regulation of an adult male-specific testosterone 15α-hydroxylase. *J. Biol. Chem.*, **263**, 17995-18002.

Matsunaga, T., Nomoto, M., Kozak, C.A. and Gonzalez, F.J. (1990) Structure and in vitro transcription of the rat CYP2A1 and CYP2A2 genes and regional localization of the CYP2A genes subfamily on mouse chromosome 7. *Biochemistry*, **29**, 1329-1341.

Maurice, M., Emiliani, S., Dalel-beluche, I., Derancourt, J. and Lange, R. (1991) Isolation and characterization of a cytochrome P450 of the IIA subfamily from human liver microsomes. *Eur. J. Biochem.*, **200**, 511-517.

Meyer, U.A., Zanger, U.M. Grant, D. and Blum, M. (1990) The genetic polymorphism of drug metabolism. *Adv. Drug Res.*, **19**, 198-206.

Meyer, U.A. (1994) Pharmacogenetics: The slow, the rapid, and the ultrarapid. *Proc. Natl. Acad. Sci. USA*, **91**, 1983-1984.

Miles, J.S., Spurr, N.K., Gough, A.C., Jowett, T., Mclaren, A.W., Brook, J.D. and Wolf, C.R. (1988) A novel human cytochrome P450 gene (P450IIB): chromosomal localization and evidence for alternative splicing. *Nucl. Acids Res.*, **16**, 5783-5795.

Miles, J.S., Bickmore, W., Brook, J.D., Mclaren, A.W., Meehan, R. and Wolf, C.R. (1989a) Close linkage of the human cytochrome P450IIA and P450IIB gene families: implications for the assignment of substrate specificity. *Nucl. Acids Res.*, **17**, 2907-2917.

Miles, J.S., Mclaren, A.W. and Wolf, C.R. (1989b) Alternative splicing in the human cytochrome P450IIB6 gene generates a high level of aberrant messages. *Nucl. Acids Res.*, **17**, 8241-8255.

Miles, J.S., Mclaren, A.W., Forrester, L.M., Glancey, M.J., Lang, M.A. and Wolf, C.R. (1990) Identification of the human liver cytochrome P450 responsible for coumarin 7-hydroxylase activity. *Biochem. J.*, **267**, 365-371.

Miles, J.S. and Wolf, C.R. (1991) Developments and perspectives on the role of cytochrome P450s in chemical carcinogenesis. *Carcinogenesis*, **12**, 2195-2199.

Molowa, D.T., Chen, W.S., Cimis, G.M. and Tan, C.P. (1992) Transcriptional regulation of the human cholesterol 7α-hydroxylase gene. *Biochemistry*, **31**, 2539-2544.

Morais, S.M.F., Wilkinson, G.R., Blaisdell, J. Nakamura, K., Meyer, U.A. and Goldstein, J.A. (1994) The major genetic defect responsible for the polymorphism of s-mephenytoin metabolism in humans. *J. Biol. Chem.*, **269**, 15419-15422.

Morishima, N., Yoshioka, H., Higashi, Y., Sogawa, K. and Fujii-Kuriyama, Y. (1987) Gene structure of cytochrome P-450 (M-1) specifically expressed in male rat liver. *Biochemistry*, **26**, 8279-8285.

Müller, G.C. and Miller, J.A. (1953) The metabolism of methylated aminoazo dyes: II. Oxidative demethylation by rat liver homogenates. *J. Biol. Chem.*, **202**, 579-587.

Nakamura, K., Goto, F., Ray, W.A., McAllister, C.B., Jacqz, E., Wilkinson, G.R. and Branch, R.A. (1985) Inter ethnic differences in genetic polymorphism of debrisoquine and mephenytoin hydroxylation between Japanese men and Caucasian populations. *Clin. Pharmacol. Ther.*, **38**, 402-408.

Nagata, K., Matsunaga, T., Gillette, J., Gelboin, H.V. and Gonzalez, F.J. (1987) Rat testosterone 7α-hydroxylase: isolation, sequence, and expression of cDNA and its developmental regulation and induction by 3-methylcholanthrene. *J. Biol. Chem.*, **262**, 2787-2793.

Narton, P.A. (1994) Alternative pre-mRNA splicing: factors involved in splice site selection. *J. Cell Sci.*, **107**, 1-7.

Nebert, D.W. and Gonzalez, F.J. (1987) P450 genes: structure, evolution, and regulation. *Annu. Rev. Biochem.*, **56**, 945-993.

Nebert, D.W., Adsnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B., Levin, W., Phillips, I.R., Sato, R. and Waterman, M.R. (1987) The P450 gene superfamily: recommended nomenclature. *DNA*, **6**, 1-11.

Nebert, D.W., Nelson, D.R., Adsnik, M., Coon, M.J., Estabrook, R.W., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Kemper, B., Levin, W., Phillips, I.R., Sato, R. and Waterman, M.R. (1989a) The P450 gene superfamily: update on listing of all genes and recommended nomenclature of the chromosomal loci. *DNA*, **8**, 1-13.

Nebert, D.W., Nelson, D.R. and Feyereisen, R. (1989b) Evolution of the cytochrome P450 genes. *Xenobiotica*, **19**, 1149-1160.

Nebert, D.W., Nelson, D.R., Coon, M.J., Estabrook, R.W., Feyereisen, R., Fujii-Kuriyama, Y., Gonzalez, F.J., Guengerich, F.P., Gunsalus, I.C., Johnson, E.F., Loper, J. C., Sato, R., Waterman, M.R. and Waxman D.J. (1991) The P450 superfamily: update on new sequences, gene mapping, and recommended nomenclature. *DNA and Cell Biology*, **10**, 1-14.

Nelson, D.R. and Strobel, H.W. (1987) Evolution of cytochrome P-450 proteins. *Mol. Biol. Evol.*, **4**, 572-593.

Nelson, D.R. and Strobel, H.W. (1988) On the membrane topology of vertebrate cytochrome P-450 proteins. *J. Biol. Chem.*, **263**, 6038-6050.

Nelson, D R., Kamataki, T., Waxman, D.J., Guengerich, F.P., Estabrook, R.W., Feyereisen, R., Gonzalez, F.J., Coon, M., Gunsalus, I.C., Gotoh, O., Okuda, K. and Nebert, D.W. (1993) The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA and Cell Biology,* **12**, 1-51.

Nouso, K., Thorgeirsson, S. and Battula, N. (1992) Stable expression of human cytochrome P450IIE1 in mammalian cells: metabolic activation of nitrosodimethylamine and formation of adducts with cellular DNA. *Cancer Res.*, **52**, 1796-1800.

Okino, S.T., Ouattrochi, L.C., Pendurthi, U.R., Mcbride, O.W. and Tukey, R.M. (1987) Characterization of multiple human cytochrome P-450 I cDNAs: the chromosomal localization of the gene and evidence for alternative RNA splicing. *J. Biol. Chem.*, **262**, 16072-16079.

Omiecinski, C.J., Walz, F.G., Jr. and Vlasuk, G.B. (1985) Phenobarbital induction of rat liver cytochrome P-450b and P-450e: quantitation of specific RNAs by hybridization to synthetic oligodeoxyribonuceotide probes. *J. Biol. Chem.*, **260**, 3247-3250.

Omiecinski, C.J. (1986) Tissue-specific expression of rat mRNAs homologous to cytochromes P-459b and P-450e. *Nucleic Acids Res.*, **14**, 1525-1539.

Omura, T. and Sato, R. (1962) A new cytochrome in liver microsomes. *J. Biol. Chem.*, **237**, 1375-1376.

Omura, T. and Sato, R. (1964) The carbon monoxide-binding pigment of liver microsomes: I. Evidence for its hemoprotein nature. *J. Biol. Chem.*, **239**, 2370-2378.

Oprian D.D. Goesky, L.D. and Coon, M.J. (1983) Properties of the oxygenated form of liver microsomal cytochrome P-450. *J. Biol. Chem.*, **258**, 8684-8691.

Osborne, T.F., Gil, G., Goldstein, J.L. and Brown, M. (1988) Operator constitutive mutation of 3-hydroxyl-3-methylglutaryl coenzyme A reductase promoter abolishes protein binding to sterol regulation element. *J. Biol. Chem.*, **263**, 3380-3387.

Palmer, G. and Reedijk, J. (1989) Nomenclature Committee of the International Union of Biochemistry (NC-IUB) nomenclature of electron-transfer protein. Recommendations. *Biochim. Biophys. Acta*, **1060**, 599-611.

Patel, D.K., Shockcor, J.P., Ghang, S.Y., Sigel, C.W. and Huber, B.E. (1991) Metabolism of a novel antitumor agent, crisnatol, by a human hepatoma cell line, HepG2, and hepatic microsomes. *Biochemical Pharmacology*, **42**, 337-346.

Pearce, R., Greenway, D. and Parkinson, A. (1992) Species differences and inter-individual variation in liver microsomal cytochrome P450 2A enzymes: effects on coumarin, dicoumarol, and testosterone oxidation. *Arch. Biochem. Biophys.*, **2998**, 211-225.

Peers, F., Bosch, X., Kaldor, J., Linsell, A. and Pluumen, M. (1987) Aflatoxin exposure, hepatitis B virus infection and liver cancer in Swaziland. *Int. J. Cancer*, **39**, 545-553.

Pelkonen, O., Raunio, H., Rautio, A., Mäenpää, J. and Lang, M.A. (1993) Coumarin 7-hydroxylase: characteristics and regulation in mouse and human. *J. Irish Colleges Physicians Surgeons*, **22**, 24-28.

Pendurithi, U.R., Lamb, J.G., Nguyen, N., Johnson, E.F. and Tukey, R.H. (1990) Characterization of the CYP2C5 gene in 21L III/J rabbits: Allelic variation affects the expression of P450IIC5. *J. Biol. Chem.*, **265**, 14662-14668.

Porter, T.D. and Coon, M.J. (1991) Cytochrome P-450: Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J. Biol. Chem.*, **266**, 13469-13472.

Poulos, T. (1991) Modelling of mammalian P450s on basis of $P450_{cam}$ x-ray structure. In Waterman, M.R. and Johnson, E.F. (eds) Methods in Enzymology. Academic press, Inc. Vol. **206**, 11-30.

Promega Protocols and Applications Guide (1991) Second Edition, Promega Corporation.

Raza, H. and Levine, W.G. (1986) Azoreduction of N,N-dimethyl-4-aminoazobenzene (DAB) by rat hepatic microsomes. Selective induction by clofibrate. *Drug Metab. Dispos.*, **14**, 19-24.

Raunio, H., Syngelma, T., Pasanen, M., Juvonen, R., Honkakoski, P., Kairaluoma, M.A., Sotaniemi, E., Lange, M.A. and Pelkonen, O. (1988) Immunochemical and catalytical studies on hepatic coumarin 7-hydroxylase in man, rat, and mouse. *Biochem. Pharmacol.*, **37** 3889-3895.

Rein, H. and Jung, C. (1993) Metabolic reactions: mechanisms of substrate oxygenation. In Schenkman, J.B. and Greim, H. (eds) Cytochrome P450, Handbook of Experimental Pharmacology, Springer-Verlag, Vol. **105**, 105-122.

Robertson, I.G.C., Serabjit-Singh, C., Croft, J.E. and Philpot, R.M. (1983) The relationship between increase in the hepatic content of cytochrome P-450, form 5, and in the metabolism of aromatic amines to mutagenic products following treatment of rabbits with phenobarbital. *Mol. Pharmacol.*, **24**, 156-162.

Ruettinger, R.T., Wen, I.-P. and Fulco, A.J. (1989) Coding nucleotides, 5' regulatory, and deduced amino acid sequences of P-450$_{BM-3}$, a single peptide cytochrome P-450: NADPH-P450 reductase from Bacillus megaterium. *J. Biol. Chem.*, **264**, 10987-10995.

Ryan, D.E., Thomas, P.E., Reik, L.M. and Levin, W. (1982) Purification, characterization, and regulation of five rat hepatic microsomal cytochrome P-450 isozymes. *Xenobiotica*, **11**, 727-744.

Ryan, D.E., Shinji, I., Wood, A.W., Thomas, P.E., Lieber, C.S. and Levin, W. (1984) Characterization of three highly purified cytochromes P-450 from hepatic microsomes of adult male rats. *J. Biol. Chem.*, **259**, 1239-1250.

Sakaguchi, M., Mihara, K. and Sato, R. (1987) A short amino-terminal segment of microsomal cytochrome P-450 functions both as an insertion signal and as a stop-transfer sequence. *EMBO J.*, **6**, 2425-2431.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Schenkman, J.B. (1993) Historical background and description of the cytochrome P450 monooxygenase system. In Schenkman, J.B. and Greim, H. (eds) Cytochrome P450, Handbook of Experimental Pharmacology, Vol. **105**, Springer-Verlag, 3-13.

Schenkman, J.B., Sligar, S.G. and Cinti, D.L. (1982) Substrate interaction with cytochrome P450. In: Schenkman, J.B. and Kupfer, D. (eds) Hepatic cytochrome P450 monooxygenase system. Pergamon, New York, 587-615.

Schüle, R. Muller, M., Kaltschmidt, C. and Renkawitz, R. (1988a) Many transcription factors interact synergistically with steroid receptors. *Science,* **242**, 1418-1420.

Schüle, R., Muller, M., Otsuka-Murakami, H. and Renkawitz, R. (1988b) Cooperativity of the glucocorticoid receptor and the CACCC-box binding factor. *Nature,* **332**, 87-90.

Shapiro, M.B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.*, **17**, 7155-7174.

Shaw, G-C. and Fulco, A.J. (1993) Inhibition by barbiturates of the binding of Bam31R repressor to its operator site on the barbiturate-inducible cytochrome $P450_{BM-3}$ gene of *Bacillus megaterium. J. Biol. Chem.*, **268**, 2997-3004.

Simmons, D.L., McQuiddy, P. and Kasper, C.B. (1987) Induction of the hepatic mixed-function oxidase system by synthetic glucocorticoids: transcriptional and post-transcriptional regulation. *J. Biol. Chem.*, **262**, 326-332.

Skoda, R., Gonzalez, F.J., Demierre, A. and Meyer, U.A. (1988) Two mutant alleles of the human cytochrome P-450db1 gene (P450C2D1) associated with genetically deficient metabolism of debrisoquine and other drugs. *Proc. Natl. Acad. Sci. USA,* **85**, 5240-5243.

Sladek, F.M. and Darnell, J.E., Jr (1992) Mechanisms of liver-specific gene expression. *Current Opinion in Genetics and Development,* **2**, 256-259.

Sladek, F.M., Zhong, W., Lai, E. and Darnell, J.E., Jr. (1990) Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Gene & Dev.*, **4**, 2353-2365.

Smith, J.R., Osborne, T.F., Brown, M.S., Goldstein, J.L. and Gil, G. (1988) Multiple steroid regulatory elements in the promoter for hamster 3-hydroxyl-3-methylglutaryl-coenzyme A synthase. *J. Biol. Chem.*, **263**,18480-18487.

Smith, T.J., Guo, Z.Y., Gonzalez, F.J., Guengerich, F.P., Stoner, G.D. and Yang, C.S. (1992) Metabolism of 4-(methylnitrosamino)-1-(3-pyridyl) -1-butanone in human lung and liver microsomes and cytochromes P-450 expressed in hepatoma cells. *Cancer Res.*, **52**, 1757-1763.

Smith, C.A.D., Gough, A., Leigh, P.N., Summers, B.A., Harding, A.H., Maranganore, D.M., Syurman, S.G., Schapira, A.H.V., Willams, A.C., Spurr, N. and Wolf, C.R. (1992) Debrisoquine hydroxylase gene polymorphism and susceptibility to Parkinson's disease. *Lancet*, **339**, 1375-1377.

Smith, G., Henderson, C.J., Parker, M.G., White, R., Bars, R.G. and Wolf, C.R. (1993) 1,4-Bis[2-(3,5-dichloropyridyloxy)]benzene, an extremely potent modulator of mouse hepatic cytochrome P-450 gene expression. *Biochem. J.*, **289**, 807-813.

Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, **98**, 503-517.

Squires, E.J. and Negishi, M. (1988) Reciprocal regulation of sex-dependent expression of testosterone 15α-hydroxylase ($P\text{-}450_{15\alpha}$) in liver and kidney of male mice by androgen for a single gene. *J. Biol. Chem.*, **263**, 4166-4171.

Strähle, U., Boshart, M., Klock, G., Stewart, F. and Schütz, G. (1989) Glucocorticoid- and progesterone-specific effects are determined by differential expression of the respective hormone receptors. *Nature*, **339**, 629-632.

Strobel, H.W., Stralka, D.J., Hammond, D.K. and White, T. (1993) Extrahepatic microsomal forms: Gastrointestinal cytochrome P450, assessment and evaluation. In Schenkman, J.B. and Greim, H. (eds) Cytochrome P450, Handbook of Experimental Pharmacology, Vol. **105**, Springer-Verlag, 363-371.

Swinney, S.W., Ryan, D.E., Thomas P.E. and Levin, W. (1987) Regioselective progesterone hydroxylation catalyzed by eleven rat hepatic cytochrome P-450 isozymes. *Biochemistry*, **26**, 7073-7083.

Taylor-Robinson, D. (1978) Cultural and serologic procedures for mycoplasmas in tissue culture. In: McGarrity, G., Murphy, D.G. & Nichols, W.W. (eds) Mycoplasma infection of cell cultures. Plenum press, London, 47-56.

Tiano, H.F., Hosokawa, M., Chilada, P.C., Smith, P.B., Wang, R-L., Gonzalez, F.J., Crespi, C.L. and Langenbach, R. (1993) Retroviral mediated expression of human cytochrome P450 2A6 in C3H/10T1/2 cells confers transformability by 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK). *Carcinogenesis*, **14**, 1421-1427.

Towin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of protein from polyacrylamide gel to nitrocellulose sheet: procedures and some applications. *Proc. Natl. Acad. Sci. USA*, **76**, 4350-4354.

Ueno, T. and Gonzalez, F.J. (1990) Complete sequence of the rat CYP2A3 gene specifically transcribed in lung. *Nucleic Acids Res.*, **18**, 4623-4624.

Ueno, T. and Gonzalez, F.J. (1990) Transcriptional control of the rat hepatic CYP2E1 gene. *Mol. Cell. Biol.*, **10**, 4495-4505.

Venepally, P., Chen, D., and Kemper, B. (1992) Transcriptional regulatory elements for basal expression of cytochrome P450IIC genes. *J. Biol. Chem.*, **267**, 17333-17338.

Vesselinovitch, S.D., Mihailovich, N., Wogan, G.N., Lombard, L.S. and Rao, K.V.N. (1972) Aflatoxin B_1, a hepatocarcinogen in the infant mouse. *Cancer Res.*, **32**, 2289-2291.

Walters, M. and Martin, D.I.K. (1992) Functional erythroid promoters created by interaction of the transcription factor GATA-1 with CACCC and AP-1/NFE-2 elements. *Proc. Natl. Acad. Sci. USA*, **89**, 10444-10448.

Wen, I-P. and Fulco, A.J. (1987) Cloning of the gene encoding a catalytically self-sufficient cytochrome P-450 fatty acid monooxygenase induced by barbiturates in *Bacillus megaterium* and its functional expression and regulation in heterologous (Escherichia coli) and homologous (*Bacillus megaterium*) hosts. *J. Biol. Chem.*, **262**, 6676-6682.

White, R.E. and Coon, M.J. (1980) Oxygen activation by cytochrome P-450. *Annu. Rev. Biochem.*, **49**, 315-356.

White, P.C., Grossberger, D., Onufer, B.J., Chaplin, D.D., New, M.I., Dupont, B. and Strominger, J.L. (1985) Two genes encoding steroid 21-hydroxylase are located near the genes encoding the fourth component of complement in man. *Proc. Natl. Acad. Sci. USA*, **82**, 1089-1093.

White, T.B., Hammond, D.K., Vasquez, H., Strobel, H.W. (1991) Expression of two cytochromes P450 involved in carcinogen activation in a human colon cell line. *Mol. Cell Biochem.*, **102**, 61-69.

Wogan, G.N. (1973a) Assessment of exposure of aflatoxins. In: Host Environment interactions in the etiology of cancer in man. IARC Monographys on Cancer Research, Vol **7**, 237-241.

Wogan, G.N. (1973b) Aflatoxin carcinogenesis. In: Hush, H. (ed.), Methods in Cancer Research, New York: Academic Press, 309-344.

Wolf, C.R. (1986) Cytochrome P-450s: polymorphic multigene families involved in carcinogen activation. *Trends in Genetics*, **2**, 209-214.

Wolf, C.R., Miles, J.S., Seilman, S., Burke, M.D., Rospendowski, B.N., Kelly, K. and Smith, W.E. (1988) Evidence that catalytic differences of two structurally homologous forms of cytochrome P-450 related to their heme environment. Biochemistry, **27**, 1597-1603.

Wolf, C.R., Smith, C.A.D., Gough, A.C., Moss, J.E., Vallis, K.A., Howard, G., Carey, F.J., Mills, K., McNee, W., Carmichael, J. and Spurr, N.K. (1992) Relationship between the polymorphisms in debrisoquine hydroxylase and cancer susceptibility. *Carcinogenesis*, **13**, 1035-1038.

Wood, A.W. and Conney, A.H. (1974) Genetic variation in coumarin hydroxylase activity in the mouse (Mus musculus). *Science*, **185**, 612-614.

Wood, A.W., Ryan, D.E., Thomas P.E. and Levin, W. (1983) Regio- and stereoselective metabolism of two C_{19} steroids by five highly purified and reconstituted rat hepatic cytochrome P-450 isozymes. *J. Biol. Chem.*, **258**, 8839-8847.

Wrighton, S.A. and Stevens, J.C. (1992) The human hepatic cytochromes P450 involved in drug metabolism. *Critical Reviews in Toxicology,* **22**, 1-21.

Yamano, S., Tatsuno, J. and Gonzalez, F.J. (1990) The CYP2A3 gene product catalyzes coumarin 7-hydroxylation in human liver microsomes. *Biochemistry*, **29**, 1322-1329.

Yamazaki, H., Inui, Y., Yun, C-H., Guengerich, F.P. and Shimada, T. (1992) Cytochrome P450 2E1 and 2A6 enzymes as major catalysts for metabolic activation of N-nitrosodialkylamines and tobacco-related nitrosamines in human liver microsomes. *Carcinogenesis*, **13**, 1789-1794.

Yamazoe, Y. and Kato, R. (1993) Activation of chemical carcinogens. In: Omura, T., Ishimura, Y. and Fujii-Kuriyama, Y. (ed.), Cytochrome P-450 (Second Edition), Kodansha Ltd., Tokyo, 159-170.

Zanger, U.M., VIlbois, F., Hardwick, J.P., and Meyer, U.A. (1988) Absence of hepatic cytochrome P450bufI causes genetically deficient debrisoquine metabolism in man. *Biochemistry*, **27**, 5447-5454.

Yano, M., Falvey, E. and Gonzalez, F.J. (1992) Role of the liver-enriched transcription factor DBP in expression of the cytochrome P450 CYP2D6 gene. *Mol. Cell. Biol.*, **12**, 2847-2854.

Ying, C.S. and Lu, A.Y.H. (1987) The diversity of substrates for cytochrome P-450. In: Guengerich, F.P. (ed) Mammalian Cytochrome P-450. CRC Press, Boca Raton, FL., Vol. **II**, 1-18.

Yu, C-Y., Motamed, K., Chen, J., Bailey, A.D. and Shen, C-K.J. (1991) The CACC box upstream of human embryonic epsilon globin gene binds Sp1 and is a functional promoter element *in vitro* and *in vivo*. *J. Biol. Chem.*, **266**, 8907-8915.

Yun, C-H., Shimada, T. and Guengerich, P. (1991) Purification and characterization of human liver microsomal cytochrome P-450 2A6. *Mol. Pharmacology*, **40**, 679-685.

Zanger, U.M., VIlbois, F., Hardwick, J.P., and Meyer, U.A. (1988) Absence of hepatic cytochrome P450bufI causes genetically deficient debrisoquine metabolism in man. *Biochemistry*, **27**, 5447-5454.

Zanussi, S., Doliana, R., Segat, D., Bonaldo, P. and Colombatti, A. (1992) The human type VI collagen gene: mRNA and protein variants of the α3 chain generated by alternative splicing of an additional 5-end exon. *J. Biol. Chem.*, **267**, 24082-24089.

Zaphiropoulos, P.G., Mode, A., Strom, A., Moller, C. Fernandez, C. and Gustafsson, J-A. (1988) cDNA cloning, sequence, and regulation of a major female-specific and growth hormone-inducible rat liver cytochrome P-450 active in 15ß-hydroxylation of steroid sulphates. *Proc. Natl. Acad. Sci. USA*, **85**, 4214-4217.

Zaphiropoulos, P.G. (1993) Differential expression of cytochrome P450 2C24 transcripts in rat kidney and prostate: evidence indicative of alternative and possibly trans splicing events. *Biochem. Biophys. Res. Commun.*, **192**, 778-786.

Zhao, J., Chan, G., Govind, S., Bell, P. and Kemper, B. (1990) Structure of 5' region and expression of phenobarbital-inducible rabbit cytochrome P450IIC genes. *DNA Cell. Biol.*, **9**, 37-48.

Zimniak, P. and Waxman, D.J. (1993) Liver cytochrome P450 metabolism of endogenous steroid hormones, bile acids, and fatty acids. In Schenkman, J.B. and Greim, H. (eds) Cytochrome P450, Handbook of Experimental Pharmacology, Vol. **105**, Springer-Verlag, 123-144.

APPENDIX 1: SOURCES OF MATERIALS

Hydroquinone
Aldrich Fine Chemicals Ltd., The Old Brickyard, New Road, Gillingham, Dorset, SP8 4JL

^{125}I-Protein A, ^{32}P-α-dCTP, ^{14}C-chloramphenicol, ECL Western blotting detection reagents, Hybond N membrane.
Amersham International plc, UK Sales Office, Lincoln Place, Green End, Aylesbury, Buckinghamshire

Chemicals
B D H, Macfarlane Robson Ltd., Burnfield Avenue, Thornliebank, Glasgow, G46 7TP
Klenow fragment, dNTP, RNase A, restriction enzymes
Boehringer Mannheim, Boehringer Mannheim House, Bell Lane, Lewes, East Sussex, BN7 1LG

Agar, bactotryptone, yeast extract, trypsin
Difco Ltd., PO Box 14B, Central Avenue, West Molesey, Surrey

En^{3}hance spray
Du Pont (UK) Ltd., Wedgewood Way, Stevenage, Hertfordshire, SG1 4YH

Cell culture media, foetal calf serum, penicillin-steptomycin, phenol, DNA 1kb ladder, guanidinium hydrochloride, guanidinium isothiocyanate, restriction enzymes
Gibco-BRL Ltd., PO Box 35, Trident House, Renfrew Road, Paisley

X-Omat AR5 X-ray film
Kodak Ltd., Box 33, Swallowdale Lane, Hemel Hempstead, Hertfordshire, HP2 7EU

Whatman 3MM paper, DE81 paper
Mackay and Lynn Ltd., 2, West Bryson Road, Edinburgh, EH11 1EH

Phosphate buffered saline
Oxoid, Wade Road, Basingstoke, Hampshire

Ficoll, dextran sulphate, hexadeoxyribonucleotides
Pharmacia Biotech, 23, Grosvenor Road, St. Albans, Herts, AL1 3AW

Anti-mouse HRP, anti-rabbit HRP
Scottish Antibody Production Unit, Glasgow and West Scotland Blood Transfusion Service, Law Hospital, Carluke, Lanarkshire, ML8 5ES

Ethidium bromide, agarose, lysozyme, MOPS, methylene blue, L-glutamine, dithiothreitol, diethyepyrocarbonate, bovine albumin, ß-mercaptoethanol, Coommassie brilliant blue R, TEMED, 4-chloro-1-naphthoflavone, dexamethasone, Folin & Coicalteau's phenol reagent, Tween 20, formamide
Sigma Chemical Co., Ltd., Fancy Road, Poole, Dorset

APPENDIX 2: PUBLICATION

Expression and alternative splicing of the cytochrome *P*-450 CYP2A7

Shaohong DING,* Brian G. LAKE,† Thomas FRIEDBERG* and C. Roland WOLF*‡

*Imperial Cancer Research Fund, Molecular Pharmacology Unit, Biomedical Research Centre, Ninewells Hospital and Medical School, Dundee DD1 9SY, Scotland, and †BIBRA Toxicology International, Woodmansterne Road, Carshalton, Surrey, SM5 4DS, U.K.

In order to investigate the relative levels of expression of human cytochrome *P*-450 (*P*-450) CYP2A genes and determine how this relates to polymorphism in coumarin hydroxylase activity, cDNA clones for members of the *CYP2A* gene family were isolated. These clones were *CYP2A6*, *CYP2A7* and an alternatively spliced version of *CYP2A7* (*CYP2A7AS*). The latter clone was missing exon 2, but contained a 10 bp segment of intron 1. Translation of *CYP2A7AS* resulted in an in-frame deletion of 51 amino acids. The expression of these cDNAs in COS-7 cells showed that both *CYP2A6* and *CYP2A7* generated a protein of molecular mass 49 kDa, whereas the protein product of *CYP2A7AS* was about 44 kDa. Only the CYP2A6 had coumarin hydroxylase activity. The relative level of CYP2A7 and CYP2A7AS mRNA was investigated by reverse transcription followed by PCR (RT-PCR) using human liver RNAs and an RNA sample from a human skin fibroblast cell line. In one of five liver RNAs studied, the aberrantly spliced CYP2A7 mRNA was 3–4-fold more abundant than the normal mRNA. The other samples contained very low levels of this mRNA species. Interestingly, CYP2A7AS mRNA was the major CYP2A7 mRNA detected in the fibroblast cell line. In this case only a protein band of 44 kDa was observed by Western-blot analysis. The relative level of mRNA encoding CYP2A6 and CYP2A7 was established in seven human liver samples by RT-PCR and found to range between 1:0.5 and 1:3. These data strengthen the previous findings that alternative splicing is an important factor in determining the levels of many human *P*-450s and that this may be subject to tissue-specific effects. Whether in this case the protein product has some function remains to be determined.

INTRODUCTION

In mammals, cytochrome *P*-450 (*P*-450) enzymes play a pivotal role in the biotransformation of endogenous compounds involved in intermediary metabolism and of xenobiotic substances such as drugs, environmental toxins and carcinogens [1,2]. It is recognized that *P*-450s are not only inducible by foreign chemicals, but also subject to hormonal regulation and genetic polymorphism. As a consequence, hepatic *P*-450 levels in man are subject to marked individual differences. The importance of these differences is exemplified by studies on the genetic polymorphism at the *CYP2D6* locus. Two distinct phenotypes, namely 'extensive metabolizer' (EM) and 'poor metabolizer' (PM), have been described [3]. The ability of individuals with the EM phenotype to metabolize some chemicals is 10–200 times higher than that of individuals with the PM phenotype [4]. In some cases the distribution of phenotypes is changed in disease populations [5] and can also be a critical determinant in inter-individual differences in toxic responses to clinical drugs.

In order to be able to determine the role of *P*-450s in adverse drug reactions and chemical toxicity, the basis for individuality in gene expression needs to be established. In this regard the genes of the human CYP2A subfamily have not been extensively investigated. Two cDNAs, designated *CYP2A6* and *CYP2A7*, have been isolated. Both genes are expressed in human liver and share 96% nucleotide sequence identity and 94% identity at the amino acid level [6,7]. CYP2A6 has been found to be responsible for the metabolism of coumarin [7,8] as well as the carcinogen aflatoxin B_1 [9,10]. The functions of CYP2A7 are currently unknown.

The level of expression of CYP2A6 is highly variable within the population. Coumarin hydroxylase activity varies greatly as well (up to 144-fold). This is consistent with the difference in hepatic CYP2A6 expression [11,12]. Northern-blot analysis with *CYP2A6* as a probe identifies two mRNAs of 2.3 and 2.8 kb in human liver [8], and immunoblotting of liver microsomes also reveals two or three bands with molecular masses of 49, 51 and 55 kDa [7,8,12]. These results suggest that more than two *CYP2A* genes are expressed in this tissue.

In order to study the factors involved in the inter-individual variability in the expression levels of human *CYP2A* genes, we have studied the expression of *CYP2A* genes in human liver. We report the identification of three mRNA species encoding CYP2A6, CYP2A7 and an alternatively spliced form of CYP2A7, designated CYP2A7AS. The identification of this latter mRNA exemplifies previous reports indicating that alternative splicing is an important factor in determining cytochrome *P*-450 levels in man.

EXPERIMENTAL

Cells

A Simian-virus-40-transformed monkey kidney fibroblast cell line (COS-7) was maintained under standard cell-culture conditions in Dulbecco's modified Eagle's medium (DMEM), supplemented with 10% foetal-calf serum, glutamine (2 mM), and antibiotics (penicillin and streptomycin); the human skin fibroblast cell line was kindly donated by Dr. S. Keyse

Abbreviations used: *P*-450, cytochrome *P*-450; CYP2A7AS, an alternatively spliced form of the cytochrome *P*-450 CYP2A7; RT-PCR, reverse transcription followed by PCR; DMEM, Dulbecco's modified Eagle's medium; DMSO, dimethyl sulphoxide.

‡ To whom correspondence should be addressed.

(Biomedical Research Centre, Dundee, Scotland) [13], and the cells were cultured in DMEM supplemented with 15% foetal-calf serum.

Preparation of radioactive DNA probes

A 0.78 kb cDNA fragment (from exon 1 to exon 5) of the CYP2A7 gene was radioactively labelled with [α-^{32}P]dCTP (3000 Ci/mmol) by random primer extension [14].

Isolation of *CYP2A* genomic clone

A human acute-lymphocytic-leukaemia DNA library in an Epstein–Barr-virus-based cosmid vector cos202 (a gift from Dr. D. Kioussis, The Middlesex Hospital Medical School, London, U.K.) was screened using a 0.7 kb cDNA fragment of CYP2A6 cDNA as a probe. Hybridization and washing of replica filters were performed as described by Kioussis et al. [15]. Analyses by restriction-enzyme digestion were carried out using standard techniques [16].

Southern-blot hybridization analysis

DNA was separated on 1%-agarose gels and then transferred to a Hybond-N nylon membrane using the conditions recommended by the manufacturer (Amersham International). The membrane was hybridized with the cDNA probe spanning exons 1 to 5 of *CYP2A7* at 65 °C overnight, followed by washing the membrane in a final salt concentration of 0.1 × SSC/0.1% SDS at 65 °C (1 × SSC is 0.15 M NaCl/0.015 M sodium citrate).

DNA sequencing

The dideoxy chain-termination method with [α-^{35}S]dATP (400 Ci/mmol) was used to sequence plasmid DNA. The DNA was prepared by alkaline lysis. RNAase A was added to final concentration of 50–100 μg/ml and incubated at 37 °C overnight. The degraded RNA was removed by precipitating plasmid DNA with 20% poly(ethylene glycol) 6000/2.5 M NaCl (3:5, v/v). Sequences were compiled and analysed by using Gene Jockey software (published and distributed by BIOSOFT).

RNA preparation from human liver and tissue-culture cells

Human livers were obtained from kidney-transplant donors. Livers were stored at −70 °C within 1 h of removal. Information about the patient case histories has been described previously [8,9]. Total cellular RNA was prepared by the guanidinium isothiocyanate method [16] or a single-step method [17] from human liver and from cultured human skin fibroblast cells. The designation of the RNA samples is the same as that described in the literature. RNA concentration and purity were estimated spectrophotometrically.

Reverse transcription and PCR (RT-PCR)

Before reverse transcription, RNA was tested for integrity by ethidium bromide staining following separation on a denaturing formaldehyde gel. The reagents were mixed in 1 × PCR buffer (Promega), as described by Innis et al. [18], in a final volume of 20 μl. These were dNTPs (1 mM), $MgCl_2$(4 mM), 1 unit/μl of RNasin, 0.1 μg of oligo$(dT)_{6-8}$, 1–5 μl of total RNA sample (10 μg) and 200 units of Moloney-murine-leukaemia-virus reverse transcriptase. A negative control without RNA was carried out in all the RT-PCR reactions, and no product was found in these negative controls. The mixture was incubated for 15 min at 23 °C, 60 min at 43 °C and then transferred to a water bath at 95 °C for 6 min. After heat treatment the reaction mixture was quickly chilled on ice. An 80 μl volume of 1 × PCR buffer, containing 10–50 pmol of each primer, $MgCl_2$ (1 mM) and 2.5 units of *Taq* polymerase (Promega) was then added. Mineral oil (100 μl) was then layered on top of the solution. To amplify the full-length *CYP2A* cDNAs, oligonucleotide A (upstream primer):

5′-CATGCTGGCCTCAGGGCTGCTT-3′

and oligonucleotide B (downstream primer):

5′-GCCTTAAGGCTTCCCCCCATTCTTATACC-3′

containing an additional *Eco*RI site were used. To amplify a cDNA fragment spanning exons 1–5, oligonucleotide A and oligonucleotide C:

5′-GAAGTCCTCCAGCCCTTGCAGC-3′

were used. The thermal cycle was: (1) denaturing; 1 min at 94 °C; (2) annealing; 1 min at 58 °C; and (3) primer extension; 1 or 2 min at 72 °C. After 30–35 cycles, the reaction mixture was incubated for 8 min at 72 °C. The PCR product was purified by phenol/chloroform (1:1, v/v) and chloroform extractions. After ethanol precipitation the amplified product using oligonucleotides A and C was resuspended in 50 μl of TE buffer (10 mM Tris/1 mM EDTA, pH 8.0). Samples (10 μl) were digested for 4 h at 37 °C with 20 units of *Pst*I.

Isolation of *CYP2A7* cDNAs

Full-length *CYP2A7* and *CYP2A7AS* cDNAs were obtained by RT-PCR using oligonucleotides A and B. The amplified products were first blunted using Klenow DNA polymerase, digested with *Eco*RI and then ligated into the *Eco*RI–*Sma*I sites of pUC19 and transformed into *Escherichia coli* DH5a.

Construction of expression plasmids

The full-length *CYP2A6* cDNA was described previously by Miles et al. [8]. In order to clone the *CYP2A6*, *CYP2A7* and *CYP2A7AS* into the mammalian expression vector $pCMV_4$ [19], the subcloned cDNAs were digested with *Eco*RI, blunted, and then cut with *Hin*dIII. The fragments were then ligated into the *Sma*I–*Hin*dIII sites of $pCMV_4$ to form the expression plasmids of $pCMV_4$-CYP2A6, $pCMV_4$-CYP2A7 and $pCMV_4$-CYP2A7AS.

Transient transfection

COS-7 cells were seeded at a density of 3×10^6cells/75 cm^2 flask and incubated overnight at 37 °C. Cells were then transfected using the DEAE-dextran method [20]. Briefly, 5 μg of plasmid DNA in 100 mM NaCl/10 mM Tris/HCl, pH 7.5, was diluted to 2.0 ml with PBS (Ca^{2+}- and Mg^{2+}-free; Gibco) containing 0.5 mg/ml of DEAE-dextran (Pharmacia; molecular mass 500 kDa). Then the mixture was added to the PBS-washed cells and incubated at 37 °C for 30 min with occasional shaking. Culture medium (10 ml), supplemented with 80 μM chloroquine, was then added to the flask and the cells incubated for 2.5 h at 37 °C. The transfection mixture was aspirated off and replaced for 2.5 min with medium (10 ml) containing 10% dimethyl sulphoxide (DMSO), followed by washing once with fresh

medium; then fresh medium (15 ml) was added and the cells were cultured for 48 h before use.

Assay of coumarin 7-hydroxylase activity in transfected COS-7 cells

The transfected cells were washed with PBS and then re-fed with 5 ml of fresh serum-free medium, 25 μl of 10 mM coumarin in DMSO and 10 μl of [3-^{14}C]coumarin (13.25 μCi in 0.5 ml of DMSO) were added to the flask and the cells were then cultured for 6 h at 37 °C. After incubation the medium was collected and an equal volume of ice-cold methanol was added. 7-Hydroxycoumarin was then assayed in the medium by h.p.l.c. analysis [21].

Western-blot analysis

The transfected COS cells and cultured skin fibroblast cells were harvested by trypsin treatment and resuspended in 0.5 ml of 10 mM sodium phosphate buffer, pH 8.0, containing 2 mM $MgCl_2$, 2 mM dithiothreitol and 1 mM EDTA. Samples were lysed by sonication using an MSE Soniprep (two 5 s bursts at an amplitude of 12 μm with the sample kept on ice). The lysed samples were centrifuged at 11600 *g* for 10 min to prepare crude supernatant and pellet fractions for Western-blot analysis. Proteins were separated by SDS/PAGE [22], transferred to nitrocellulose membrane and probed with anti-(rat CYP2A) antiserum using the method of Lewis et al. [23]. For human liver microsomal preparation [8], 10 μg of protein was loaded per track, whereas for transfected COS cells and skin fibroblast cells, 100 μg of crude supernatant and pellet fractions were used.

RESULTS

Determination of CYP2A mRNA levels

Analysis of the human *CYP2A6* and *CYP2A7* cDNA sequences [7] shows the presence of a *Pst*I restriction site at bp 143 (+1 indicates the start of the open reading frame) of *CYP2A6*, which is absent in *CYP2A7*. This restriction site was used to establish the relative mRNA levels encoded by these two genes. Hepatic mRNA was reverse-transcribed, and CYP2A6 and CYP2A7 were amplified by PCR using oligonucleotides A and C as primers. The expected 750 bp fragments of *CYP2A6* and *CYP2A7* were observed (Figure 1b; the tracks indicated by a 'U'). After digestion with *Pst*I (20 units and 4 h incubation at 37 °C), *CYP2A6* cDNA was, as expected, cut into two fragments of 607 bp and 143 bp, whereas *CYP2A7* cDNA was unaffected by this procedure. During subsequent electrophoresis the 143 bp fragment migrated out of the gel and was not revealed (Figure 1b). In a separate experiment, 0.4 mg of pBluescript DNA was added into the reaction mixture as an internal control (Figure 1a) to verify a complete digestion. In order to ensure that the analysis was reproducible, the RT-PCR and *Pst*I digestion for samples L8, L6 and L4 were carried out twice. The results revealed that the ratio of *CYP2A7* to *CYP2A6* in these three RNA samples were reproducible (results not shown). Both the *CYP2A6* and *CYP2A7* cDNA were expressed in all liver samples. The ratio of the expression level of *CYP2A7* to that of *CYP2A6* was subject to some variation, ranging from slightly lower than 1:1 in sample L6, to about 3–4:1 in sample L4. The relative expression of *CYP2A7* to *CYP2A6* in one RNA sample (L6) was determined over a range of PCR cycles. The results showed that the expression level of *CYP2A7* to that of *CYP2A6* was constant with the number of PCR cycles (Figure 1c).

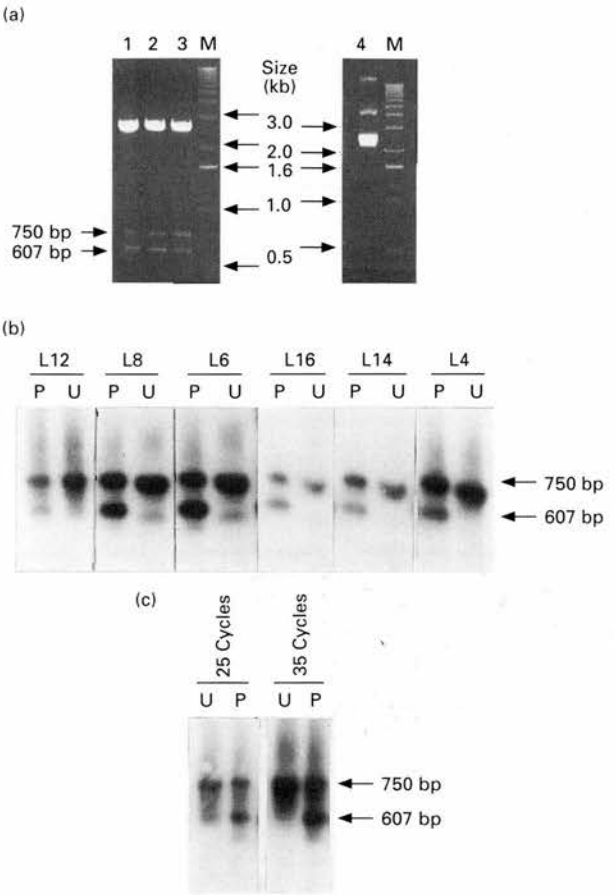


Figure 1 Relative amounts of CYP2A6 and CYP2A7 mRNA in six human livers

The amplified DNA was cut with 20 units of *Pst*I and longer incubation (4 h). (**a**) Agarose-gel electrophoresis of *Pst*I digested RT-PCR products. A 0.4 μg portion of plasmid pBluescript DNA (2.94 kb) was added to the reaction mixture to verify complete digestion. Lanes M, 1 kb ladder marker; lanes 1–3, digested pBluesript DNA (2.94 kb) with RT-PCR products from RNA samples L8, L9 and L11 respectively; 4, uncut pBluesript DNA. (**b**) Southern-blot analysis of *Pst*I digested RT-PCR products. The fragment of 607 bp represents the digested *CYP2A6*, and that of 750 bp is *CYP2A7*. (**c**) The ratio of the expression of the *CYP2A7* relative to *CYP2A6* over a range of PCR cycles with RNA sample L6; lanes U, uncut PCR product; lanes P, *Pst*I-digested product. The designation of the samples is the same as that described in the literature [8,9].

The identification of an alternatively spliced CYP2A7 mRNA

As part of the analysis of CYP2A6 and CYP2A7 mRNAs, RT-PCR was carried out on the entire coding region of the *CYP2A* genes using a human total liver RNA sample (L8) and oligonucleotide primers; A (complementary to the 5′ end, bp 1–22) and B (complementary to the 3′ end, bp 1583–1564). In addition to the expected fragment of 1.6 kb, a weak fragment of about 1.45 kb was observed. The PCR products were subcloned into the vector pUC19. *Pst*I digestion was used to screen the colonies, as this site is present at bp 143 (+1 indicates the start of the open reading frame) of CYP2A6, but absent in CYP2A7. Three colonies with different *Pst*I digestion patterns were isolated and sequenced. The results indicated that the colonies with a 1.6 kb insert contained a *CYP2A6* or a *CYP2A7* cDNA. The colony with a 1.45 kb insert appeared to contain an aberrantly spliced version of CYP2A7 (CYP2A7AS).

To determine the basis for the alternative splicing, we isolated

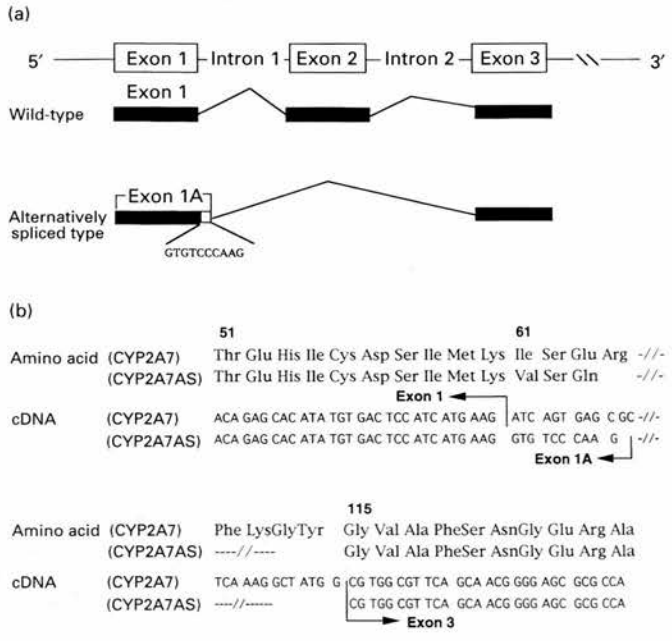


Figure 2 (a) Structures of the normal and alternatively spliced CYP2A7 mRNAs, and (b) deduced amino acid sequences of CYP2A7 and CYP2A7AS

(**a**) Exon 1A contains exon 1 plus the first 10 bp of intron 1. (**b**) In CYP2A7AS the 54 amino acids of exon 2 are replaced by the three amino acid residues Val-Ser-Gln.

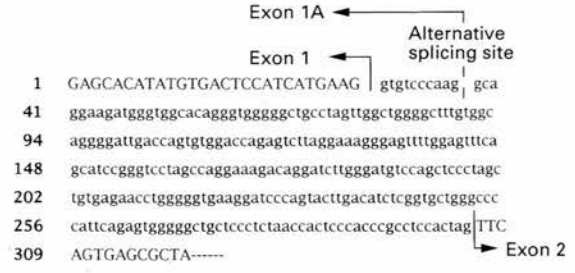


Figure 3 Sequence covering the alternatively spliced region of CYP2A7

The alternative splice site is indicated by a vertical broken line. The exon sequences are designated by upper-case letters; intron sequences are indicated by lower-case letters.

one clone from a human genomic cosmid library, designated CoIIA, using a 0.7 kb *CYP2A6* cDNA fragment (exons 1–5) as probe. The restriction-enzyme digestion followed by Southern-hybridization analysis using three *CYP2A6* cDNA probes (a 150 bp *SalI–PstI* fragment containing exon 1 [6], a 400 bp and a 300 bp PCR fragment containing exons 5 and 6, and exons 8 and 9 respectively) revealed that the CoIIA contained an entire coding region within the 35 kb insert. In order to sequence the coding region in CoIIA, several restriction fragments of CoIIA were subcloned into pUC18 vector and sequenced. In addition, oligonucleotides based on the intron/exon organization of *CYP2A* subfamily genes were used to sequence all exons and parts of the introns.

Comparing the genomic sequence of *CYP2A7* gene with those of cloned *CYP2A* cDNAs, we found that the CYP2A7AS was an aberrantly spliced version of CYP2A7 in which the 163 bp of exon 2 was replaced by a 10 bp segment (Figure 2a). The alternatively spliced exon 1 was designated exon 1A. Translation of CYP2A7AS mRNA would result in an in-frame deletion of 51 amino acids to generate a protein product of molecular mass 44 kDa. The sequence of the 10 bp segment adds three amino acid residues at position 60 and links to the amino acid at position 114 in exon 3 (Figure 2b). All the intron/exon junctions conformed to the normal GT/AG consensus splice recognition site. However, the intron 1 contained an additional 5′-splice site, G/gcagg (exon sequences are designated by upper-case letters and intron sequences by lower-case letters), which resulted in CYP2A7AS (Figure 3). This alternative 5′-splice site does not conform to the GT/AG consensus sequence; however, it is similar to a 'non-conforming' 5′ splice site sequence found in a few genes [24,25].

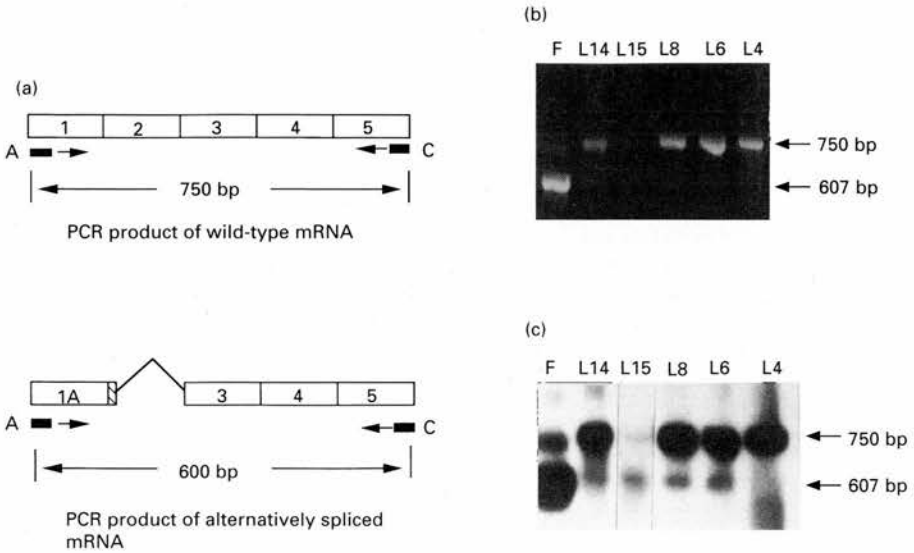
Expression of the alternatively spliced CYP2A7 in human liver and a skin fibroblast cell line

To investigate the extent of alternative splicing of CYP2A7, five human liver RNA samples and total RNA from cultured human skin fibroblast cells were assayed by RT-PCR. The expected wild-type PCR product is a fragment of 750 bp, and the product from the alternatively spliced mRNA is 600 bp (Figure 4a). An ethidium bromide-stained band of 750 bp appeared in four out of five liver RNA samples, whereas a weak band of 750 bp and a much stronger 600 bp band existed in the PCR product of the skin-fibroblast-cell RNA (Figure 4b). After transfer to a Hybond-N nylon membrane and hybridizing with a 0.78 kb probe (spanning exons 1–5 of *CYP2A7*), four of five samples, including sample L15 (after longer exposure), contained alternatively spliced CYP2A7 mRNA. Variability in the relative level of this mRNA as compared with normal CYP2A7 mRNA was found between samples (Figure 4c). Indeed, in RNA sample L15 the level of the alternatively spliced mRNA was 3–4-fold higher than the correctly spliced transcript. However, in the other samples the normal transcript was the predominant mRNA species. Interestingly, the major CYP2A7 mRNA species in the skin fibroblast cell line appeared to be the alternatively spliced mRNA, with only a very low amount of the normal transcript (Figures 4b and 4c, track F).

Functional analysis of CYP2A transcripts

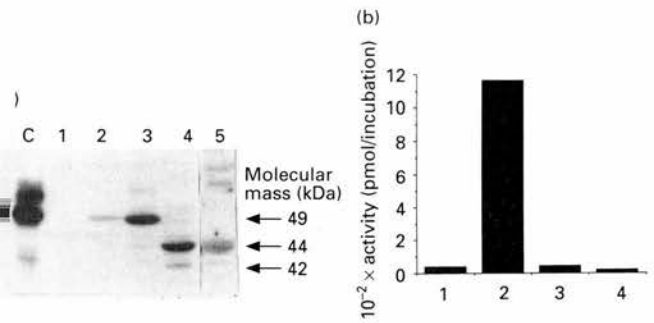
To investigate whether these mRNAs could be translated, the *CYP2A6*, *CYP2A7* and alternatively spliced *CYP2A7* cDNAs were subcloned into the vector $pCMV_4$ and transfected into COS cells. Western-blot analysis of the membrane fractions showed that all three cDNAs gave protein products (Figure 5, tracks 2–4). The molecular masses of CYP2A6 and CYP2A7 proteins were identical (49 kDa), and they showed a mobility identical with that of the major immunostained band identified in human liver microsomes with anti-(rat CYP2A1) antibody. The alternatively spliced CYP2A7 gave two protein bands (molecular masses 44 kDa and 42 kDa; Figure 5, track 4). Since the abnormal mRNA often leads to an unstable protein product, the 42 kDa protein could be a degradation product. A protein band which co-migrated with 44 kDa protein was also found in the crude membrane fraction of the human skin fibroblast cells, but not in the human liver microsomal sample. No detectable 49 kDa protein was observed in the fibroblast cell line (Figure 5a, track 5).

In order to determine whether the expressed proteins are catalytically active, coumarin hydroxylase activity in the transfected cells was measured by h.p.l.c. (Figure 5b). Relative to controls, cells expressing *CYP2A6* had considerable activity. However, no activity could be measured in cells transfected with the different forms of *CYP2A7*.



gure 4 Analysis of CYP2A7 species

Analysis procedure: RT-PCR was carried out with primers A and C as described in the Experimental section. (**b**) The products were analysed by ethidium bromide staining. Lane F, human roblast total RNA. (**c**) Southern hybridization with a 0.78 kb cDNA probe containing exons 1–5 of *CYP2A7*. The track of L15 underwent a longer exposure.



gure 5 (a) Expression of *CYP2A* cDNAs in COS-7 cells, and (b) coumarin-hydroxylase activity

cDNAs encoding CYP2A6, CYP2A7 or CYP2A7AS were subcloned into $pCMV_4$ and nsfected into COS-7 cells as described in the Experimental section. Shown is a Western-tting analysis of proteins from the crude pellet fractions of transfected COS cells (100 μg) of cultured human skin fibroblast cells (100 μg) using an antibody to rat CYP2A1. Human r microsomal protein (10 μg, lane C) was used as a positive control. Track 1, untransfected 3-7 cells; track 2, $pCMV_4$-CYP2A6; track 3, $pCMV_4$-CYP2A7; track 4, $pCMV_4$-CYP2A7AS track 5, human skin fibroblast cells (in the same Western-blotting analysis). (**b**) A proportion the cells from the same samples analysed by Western-blot analysis were assayed for umarin hydroxylase activity as described in the Experimental section. Activities are expressed pmol/6 h per incubation mixture ($= 5 \times 10^6$ cells). Tracks are: 1, untransfected cells; 2, MV_4-2A6; 3, $pCMV_4$-2A7; 4, $pCMV_4$-2A7AS.

SCUSSION

the present paper we have shown that CYP2A7 mRNA can be ernatively spliced to give a transcript missing exon 2, but ntaining three amino acids derived from intron 1. This mRNA n be translated to give a protein product of the predicted lecular mass in cDNA-directed expression and in a human roblast cell line. The relative hepatic level of the CYP2A7AS sus CYP2A7 mRNA varied considerably between individuals, and in one case CYP2A7AS was the more abundant of these two mRNA species. In the human fibroblast cell line, CYP2A transcripts were detected, but the major product was CYP2A7AS. The basis for the variability in alternative splicing of CYP2A7 mRNA is not known and could be determined by either genetic and/or environmental factors.

It is intriguing that aberrantly spliced mRNAs have also been reported for several other genes in the CYP2 family. For example, two mRNAs are derived from the rat *CYP2C6* gene by alternative splicing in exon 8. This, like many other examples, leads to a disruption of the open reading frame. Although the aberrantly spliced mRNA can be translated into a truncated protein, its haem-binding capacity is lost and therefore the protein cannot function as a *P*-450 mono-oxygenase [26]. The transcript of human *CYP2B6*, whose expression may be co-regulated with *CYP2A* [12], is also alternatively spliced, and at least four mRNA species are derived from this gene [25]. The relative levels of different mRNAs are subject to considerable individual variability. Similar to the findings here, a variant of CYP2B6 is generated using a cryptic 'non-conforming' 5′-splice site, G/gcaag. Alternative splicing has also been described for human *CYP2D* genes [4,27]. Taken together, these results suggest that alternative splicing is an important determinant in the expression of many *P*-450 genes, and this effect will contribute to the inter-individual variation in the enzyme levels. In addition, alternative splicing using cryptic non-GT-conforming 5′-splice site is considered as a rate-limiting regulation for some genes [24].

CYP2A7AS produces a truncated protein of molecular mass 44 kDa. This protein still contains the conserved *P*-450 haem-binding region and could conceivably still function as a mono-oxygenase enzyme. However, CYP2A7AS does not contain exon 2, which might form a potential transmembrane domain [28] and contains specific amino acids responsible for substrate recognition [29]. The results showed that the CYP2A7AS protein was associated with the crude membrane fraction (Figure 5a). No detectable CYP2A7AS protein was found in the crude supernatant fraction (results not shown). These results agree with

recent membrane-topology models of *P*-450, suggesting that only exon 1 of the *P*-450s codes for the membrane anchor [30]. Since the expressional level of CYP2A7AS in COS-7 cells was too low to establish whether the truncated protein still binds haem or not, studies to establish whether this is the case using other expression systems are in progress.

CYP2A6 is reported to be the major enzyme catalysing coumarin 7-hydroxylation in human liver [7,8]. This conclusion was based on the observed correlation between the level of CYP2A6 protein and the enzyme activity in human livers. Our finding that CYP2A7 had exactly the same mobility as CYP2A6 on SDS/PAGE indicated that CYP2A7 may also contribute to this activity. However, cDNA-directed expression of CYP2A6 and CYP2A7 in COS-7 cells indicated that only CYP2A6 had coumarin 7-hydroxylase activity, although the cellular expression of this protein was much lower than that of the CYP2A7 protein. The lack of correlation of the level of CYP2A6 protein with coumarin hydroxylase activity in some individual samples could be explained if CYP2A7 were the major protein present.

Enzymes in the CYP2A subfamily play a role in the metabolic activation of promutagens, such as nitrosamines [31,32], benzo[*a*]pyrene and aflatoxin B_1 [9,33–35]. To date no substrates for CYP2A7 have been identified. Analysis of mRNA levels indicates that, in certain samples, CYP2A7 is expressed at higher concentration than CYP2A6. It will therefore be important to determine the substrates for this enzyme.

We are grateful to Dr. C. J. Henderson and Ms. A. W. McLaren for some human liver RNA samples.

REFERENCES

1 Wislocki, P. G., Miwa, G. T. and Lu, A. Y. H. (1980) in Enzymatic Basis of Detoxication (Jakoby, W. B., ed.), vol. 1, pp. 135–182, Academic Press, New York
2 Wrighton, S. A. and Stevens J. C. (1992) Crit. Rev. Toxicol. **22**, 1–21
3 Eichelbaum, M., Kroemer, K. H. and Mikus, G. (1992) Toxicol. Lett. **1**, 115–122
4 Gough, A. C., Miles, J. S., Spurr, N. K., Moss, J. E., Gaedigk, A., Eichelbaum, M. and Wolf, C. R. (1990) Nature (London) **347**, 773–776
5 Smith, C. A. D., Gough, A. C., Leigh, P. N., Summers, B. A., Harding, A. E., Maranganore, D. M., Sturman, S. G., Schapira, A. H. V., Williams, A. C., Spurr, N. K. and Wolf, C. R. (1992) Lancet **339**, 1375–1377
6 Miles, J. S., Bickmore, W., Brook, J. D., McLaren, A. W., Meehan, R. M. and Wolf, C. R. (1989) Nucleic Acids Res. **17**, 2907–2917
7 Yamano, S., Tatsuno, J. and Gonzalez, F. J. (1990) Biochemistry **29**, 1322–1329
8 Miles, J. S., Mclaren, A. W., Forrester, L. M., Glancey, M. J., Lang, M. A. and Wolf, C. R. (1990) Biochem. J. **267**, 365–371
9 Forrester, L. M., Henderson, C. J., Glancey, M. J., Back, D. J., Park, B. K., Ball, S. E. Kitteringham, N. R., McLaren, A. W., Miles, J. S., Skett, P. and Wolf, C. R. (1992) Biochem. J. **281**, 359–368
10 Crespi, C. L., Penman, B. W., Steimel, D. T., Gelboin, A. V. and Gonzalez, F. J. (1991 Carcinogenesis **12**, 355–359
11 Pelkonen, O., Raunio, H., Rautio, A., Mäenpää, J. and Lang, M. A. (1993) J. Ir. Coll. Phys. Surg. **22**, 24–28 (Suppl. 1)
12 Pearce, R., Greenway, D. and Parkinson, A. (1992) Arch. Biochem. Biophys. **298**, 211–225
13 Keyse, S. M. and Emslie, E. A. (1992) Nature (London) **359**, 644–647
14 Feinberg, D. P. and Vogelstein, B. (1983) Anal. Biochem. **136**, 6–13
15 Kioussis, D., Wilson, F., Daniels, C., Leveton, C., Taverne, J. and Playfair, J. H. L. (1987) EMBO J. **6**, 355–361
16 Maniatis, T, Fritsch, E.F and Sambrook, J. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor University Press, Cold Spring Harbor, NY
17 Chomczynski, O. and Sacchi, N. (1987) Anal. Biochem. **162**, 156–159
18 Innis, M. A., Gelfand, D. H., Sninsky, J. J. and White, T. J. (1990) PCR Protocols: A Guide to Methods and Applications, pp. 21–27, Academic Press, San Diego
19 Andersson, S., Davis, D. L., Dahlbäck, H., Jörnvall, H. and Russell, D. W. (1989) J. Biol. Chem. **264**, 8222–8229
20 Cullen, B. R. (1986) Methods Enzymol. **152**, 684–704
21 Van Iersel, M. L. P. S., Henderson, C. J., Walters, D. G., Price, R. J., Wolf, C. R. and Lake, B. G. (1994) Xenobiotica, in the press
22 Laemmli, U. K. (1970) Nature (London) **227**, 680–685
23 Lewis, A. D., Hickson, T. D., Robson, C. N., Harris, A. L., Hayes, J. D., Griffiths, S. / Manson, M. M., Hall, A. E., Moss, J. E. and Wolf, C. R. (1988) Proc. Natl. Acad. Sc U.S.A. **85**, 8511–8515
24 Shapiro, M. B. and Senapathy, P. (1987) Nucleic Acids Res. **15**, 7155–7174
25 Miles, J. S., McLaren, A. W. and Wolf, C. R. (1989) Nucleic Acids Res. **17**, 8241–8255
26 Kimura, H., Sogawa, K., Sakai, Y. and Fujii-Kuriyama, Y. (1989) J. Biol. Chem. **26** 2338–2342
27 Gonzalez, F. J., Skoda, R. C., Kimura, S., Umeno, M., Zanger, U. M., Nebert, D. W., Gelboin, H. V., Hardwick, J. P. and Meyer, U. A. (1988) Nature (London) **331**, 442–446
28 Nelson, D. R. and Strobel, H. W. (1988) J. Biol. Chem. **263**, 6038–6050
29 Gotoh, O. (1992) J. Biol. Chem. **267**, 83–90
30 Black, S. D. (1992) FASEB J. **6**, 680–685
31 Yamazaki, H., Inui, Y., Yun, C-H., Guengerich, F. P. and Shimada, T. (1992) Carcinogenesis **13**, 1789–1794
32 Camus, A. M., Geneste, O., Honkakoski, P., Bereziat, J. C., Henderson, C. J., Wolf, C. R. and Lang, M. A. (1994) Carcinogenesis, in the press
33 Aoyama, T.,Yamano, S., Guzelian, P. S., Gelboin, H. V., and Gonzalez, F. J. (1991) Proc. Natl. Acad. Sci. U.S.A. **87**, 4790–4793
34 Forrester, L. M., Neal, G. E., Judah, D. J., Glancey, M. J. and Wolf, C. R. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 8306–8310
35 Tiano, H. F., Hosokawa, M., Chilada, P. C., Smith, P. B., Wang, R-L., Gonzalez, F. J Crespi, C. L. and Langenbach, R. (1993) Carcinogenesis **14**, 1421–1427

Received 17 June 1994/27 September 1994; accepted 7 October 1994